Introduction to mAdb

Esther Asaki, Yiwen He, John Powell

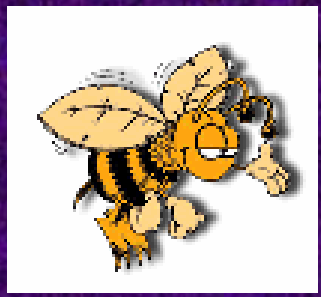
I. Introducing the mAdb system

II. Putting your data in mAdb

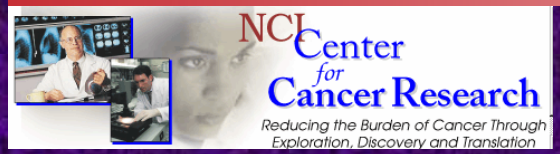
III. Evaluating array quality

IV. Getting started with analysis

V. Managing your data



March 30, 2006

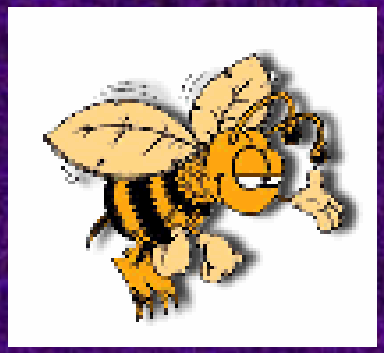


Logging into the Training Server

- Point your browser at http://madb-training.cit.nih.gov – for use in class only!
- Your username is on the card on your desk
- Today's Password is on whiteboard near door
- Don't request a mAdb account on the training server!! – request at madb.nci.nih.gov or madb.niaid.nih.gov
- Do not maximize your browser; leave room to see and click on other windows

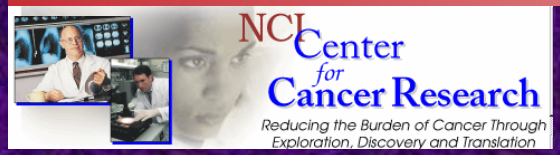
I. Introducing the mAdb system

mAdb BioInformatics Project



Goal:

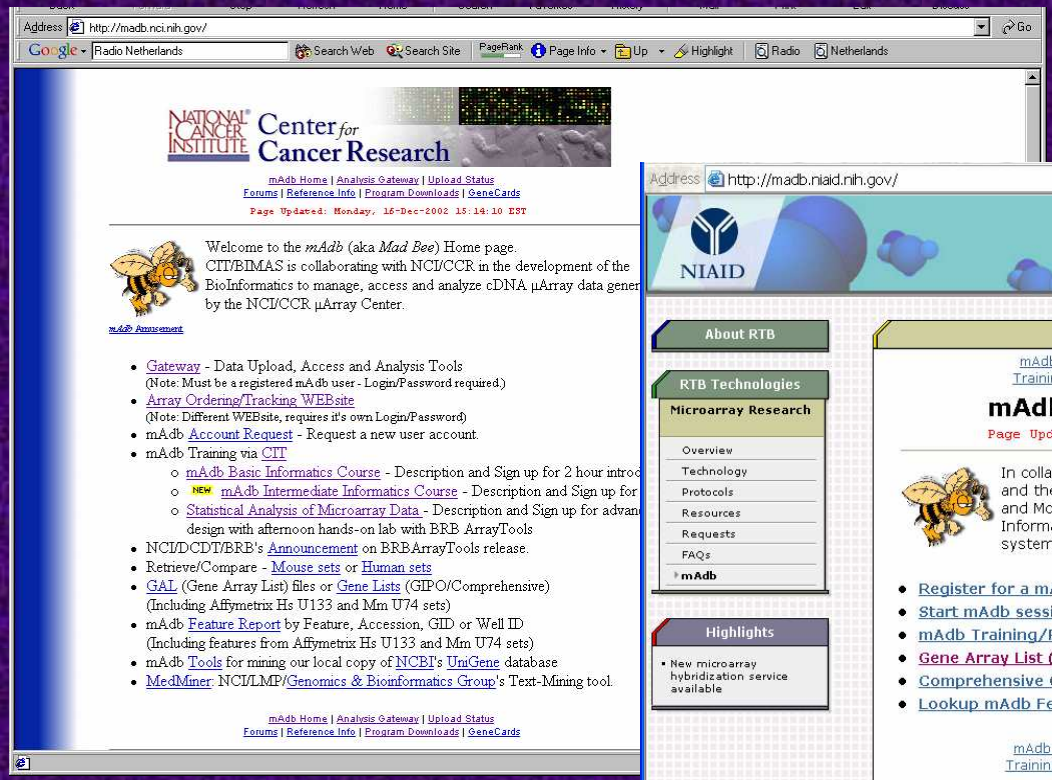
- **Provide an integrated set of web-based analysis tools and a data management system for storing and analyzing cDNA/oligo/Affy Gene Expression data using open systems design, focusing on 2 color array slides.**
- **System currently supports spotted arrays routinely produced by the NCI, NIAID, and FDA Microarray Centers**
- **Currently support Axon GenePix, Perkin-Elmer QuantArray, and Arraysuite II / IP Lab image analysis software (Yidong Chen, NHGRI) for two-color, "Pat Brown-type" spotted arrays**
- **Affymetrix now available after a consultation to learn needed parameters – limited number of chips supported right now (mouse, human, rat)**

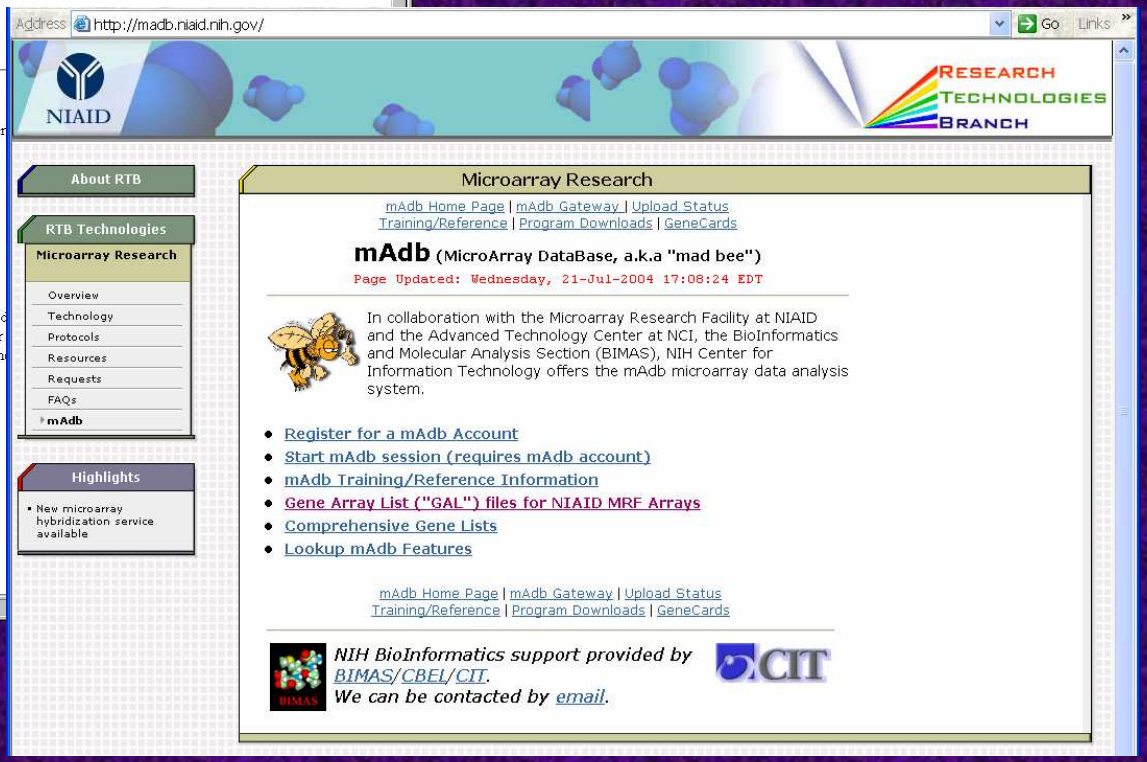


mAdb Home Page URLs

http://madb.nci.nih.gov

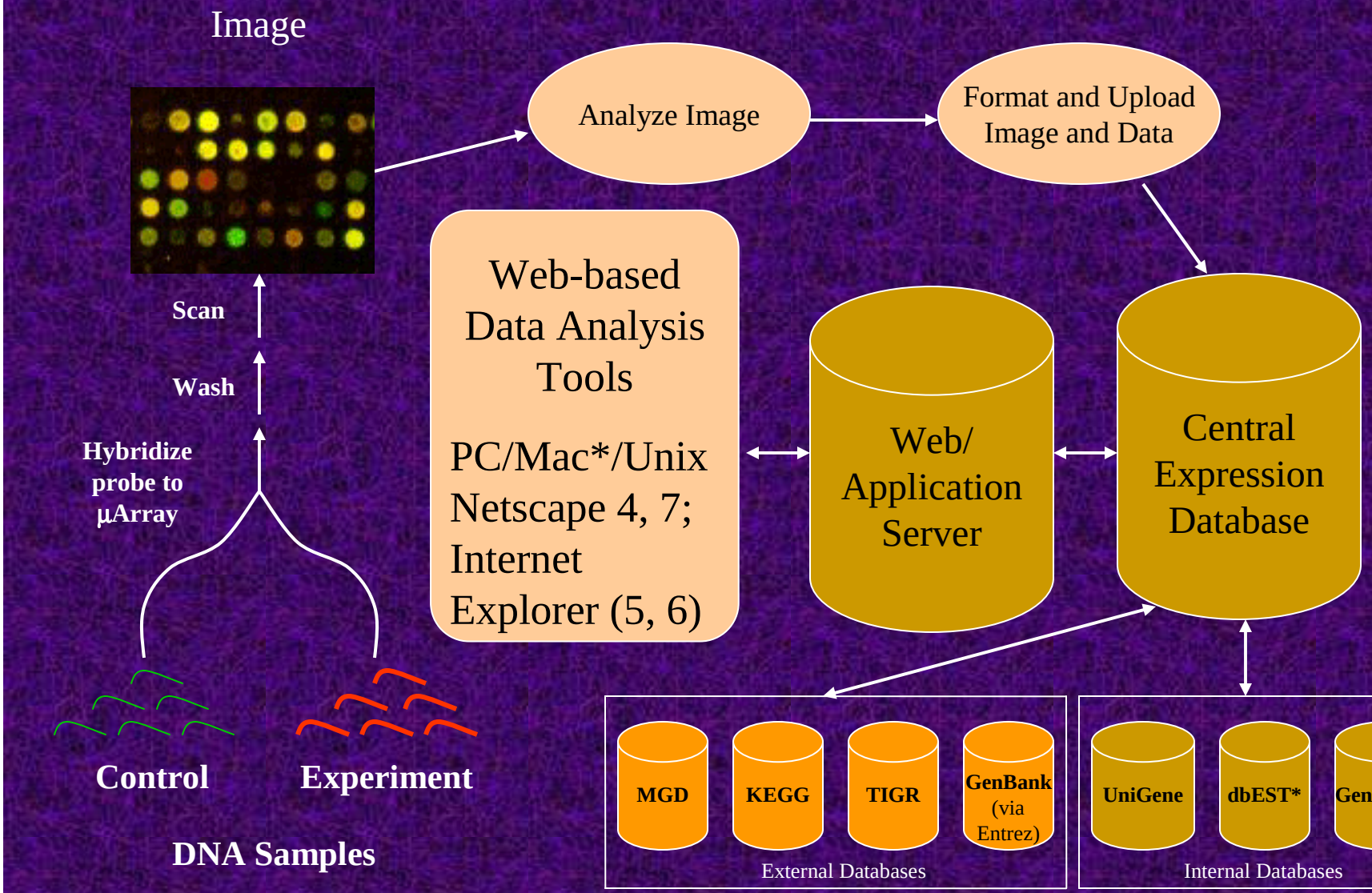
http://madb.niaid.nih.gov





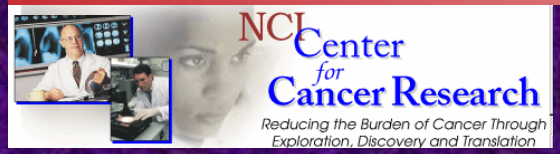
For support, please e-mail: madb_support@bimas.cit.nih.gov

Architecture for µArray Informatics



mAdb Quick Facts

- 58234 arrays uploaded since Feb. 2000 – now average ~750 per month uploaded over last year (updated 3/30/06)
- over 1,300 registered users (NIH and collaborators)
- Among the largest collections of microarray data in the world, although data sharing is determined by each investigator – no one has access to all the data
- MIAME capable format – available upon request
- Can help you get data into public repositories – GEO (NCBI), ArrayExpress (EBI)



NIAID

CIT

mAdb System Features

- Gene Discovery
  - Outlier detection – row retrieval tools
  - Scatter plots
  - *Ad hoc* keyword queries
  - Multiple array viewer
- Class Comparison
  - t-test; Wilcoxon; ANOVA; Kruskal-Wallis; SAM
- Class Prediction
  - PAM classifier
- Class Discovery (unsupervised)
  - Clustering – Hierarchical, K-means, SOMs
  - Multidimensional Scaling
  - Principal Components Analysis
- Pathway summary – GO, KEGG, BioCarta
- Boolean comparison of data

Class #412 - Analyzing Microarray Data using the mAdb System

Live Demo

Home Page Notes

- Special Notices to Users
- Analysis Gateway link
- Account Requests link
- Array Tracker link – N.B. separate login & password!
- Training signup links
- GAL/GIPO links

mAdb GAL files

Current (2002) NCI Production Gene Array List Files (GAL files) (blocks x columns x rows)

- NEW Earlier NCI production printings
- Custom printings
- NIAID printings
- NEW FDA printings
- Mini-lymphochip GAL files (restricted to registered users)

Human Array Sets	
GAL File	**Array Sets**
Hs-UniGEM2-v2px-32Bx18Cx18R.gal Generated Tuesday, 21-May-2002 09:21:59 EDT Note: Also use for 2.1px, 2.3px, 2.4px, 2.5px, 2.6px, 5.0px See below for special **3.5px** gal file NEW See below for special **4.0px** gal file NEW See below for special **4.1px** gal file NEW See below for special **4.2px** gal file	Hs-UniGEM2-v2.4p1 Hs-UniGEM2-v2.4p2 Hs-UniGEM2-v2.4p3 Hs-UniGEM2-v2.4p4 Hs-UniGEM2-v2.4p5 Hs-UniGEM2-v2.4p8 Hs-UniGEM2-v2.4p9 Hs-UniGEM2-v2.5p3 Hs-UniGEM2-v2.5p4 Hs-UniGEM2-v2.5p5 Hs-UniGEM2-v2.5p6 Hs-UniGEM2-v2.5p7 Hs-UniGEM2-v2.5p11 Hs-UniGEM2-v2.6p2 Hs-UniGEM2-v2.6p3 Hs-UniGEM2-v2.6p6 Hs-UniGEM2-v2.6p7 Hs-UniGEM2-v2.6p8 Hs-UniGEM2-v2.6p9 Hs-UniGEM2-v2.6p10 Hs-UniGEM2-v5.0p1 Hs-UniGEM2-v5.0p2 Hs-UniGEM2-v5.0p3 Hs-UniGEM2-v5.0p4 Hs-UniGEM2-v5.0p5 Hs-UniGEM2-v5.0p6 Hs-UniGEM2-v5.0p7 Hs-UniGEM2-v5.0p8 Hs-UniGEM2-v5.0p9 Hs-UniGEM2-v5.0p10 Hs-UniGEM2-v5.0p12 Hs-UniGEM2-v5.0p13 Hs-UniGEM2-v5.0p14 Hs-UniGEM2-v5.0p15 Hs-UniGEM2-v5.0p16 Hs-UniGEM2-v5.0p17 Hs-UniGEM2-v5.0p18
Hs-UniGEM2-v3.5px-32Bx19x17R.gal Generated Tuesday, 21-May-2002 09:33:10 EDT	Hs-UniGEM2-v3.5p1 Hs-UniGEM2-v3.5p2
Hs-UniGEM2-4.0px-32Bx18Cx18R.gal Generated Monday, 25-Nov-2002 15:03:35 EST	Hs-UniGEM2-v4.0p2 Hs-UniGEM2-v4.0p4 Hs-UniGEM2-v4.0p5 Hs-UniGEM2-v4.0p6 Hs-UniGEM2-v4.0p7 Hs-UniGEM2-v4.0p8 Hs-UniGEM2-v4.0p9 Hs-UniGEM2-v4.0p10 Hs-UniGEM2-v4.0p11
Hs-UniGEM2-4.1px-32Bx18Cx18R.gal Generated Monday, 25-Nov-2002 15:34:59 EST	Hs-UniGEM2-v4.1p1

- Shows the actual GAL (Gene Array list) files – link block, row, column to what DNA is spotted there
- One printset layout is usually used for many lots of slides
- Please e-mail mAdb support if you cannot find your GAL file listed

Application Program Downloads

mAdb Program Downloads

Page Updated: Friday, 15-Aug-2003 08:45:58 EDT

		Program	Description	Author	Version	Updated	Download	Manual
Axon Inc. Software This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all. **Axon's Web Site**		**GenePix Pro 5**	Fully integrated acquisition and analysis software for the GenePix 4000, 4100 & 4200. Download to a folder of your choice and then run to start the installation process.		5.0.1.13 History	8/15/2003 (Posted here 8/15/2003)	Download	Users Guide & Tutorial (PDF)
Axon Inc. Software This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all. **Axon's Web Site**		**GenePix Pro 4**	Fully integrated acquisition and analysis software for the GenePix 4000 & 4100. Download to a folder of your choice and then run to start the installation process.		4.0.1.17 History	(Posted here 3/12/2003)	Download	Manual Axon Scanner Manual (PDFs)
Axon Inc. Software This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all. **Axon's Web Site**		**GenePix Pro 3**	Fully integrated acquisition and analysis software for the GenePix 4000A. Download to a folder of your choice and then run to extract the installation files. Then run the extracted file setup.exe and follow installation instructions		3.0.6.89 History	(Posted here 02/18/2002)	Download	Manual Axon Scanner Manual (PDFs)
Stanford Genome Analysis Group Software It is available free of charge to academic and non-profit institutions. **Eisen Lab Download Site**	SA	**ScanAlyze**	Image Analysis (extracts data from fluorescence images of arrays)	Michael Eisen	2.44	11/15/99	Download	Manual (PDF)
		Cluster	Perform Hierarchical Clustering, Self-organizing Maps, k-Means Clustering, and More	Michael Eisen	2.11.01	7/10/2000 (Posted here 10/26/2000)	Download	Manual (PDF)
		TreeView	Graphical Viewing and Browsing of Cluster Results	Michael Eisen	1.5	04/2000 (Posted here	Download	
EASE: Expression Analysis Systematic Explorer Developed by the Laboratory of Immunopathogenesis and Bioinformatics, SAIC Frederick **EASE Web Site**		**EASE**	For finding "biological meaning" of gene lists via three functions: biological theme over-representation analysis, creation of annotation tables, and automated loading of genes into various online tools.	Doug Hosack	Revision history	Current version	Link to Download	Online help (Online)
MAExplorer Developed by and Available from LECB/FCRF/NCI. **MAExplorer Web Site**		**MAExplorer**	A Java data mining application for gene expression data using a variety of statistical, clustering, direct-manipulation graphical, spreadsheet and Web access methods.	Peter Lemkin	Revision History	Current version	Link to Download	Manual (Online) Use with mAdb data (PDF)

Various versions of GenePix are supported

Page accessible from NIH network only

Prefer GenePix updates obtained from this page – validated to work with mAdb

mAdb Training/Reference Page

mAdb Training/Reference Information

Page Updated: Monday, 22-Nov-2004 17:17:41 EST

- mAdb Training Classes via CIT
  - Introduction to mAdb - Description and Sign up for a 3 hour introductory class on using mAdb.
  - Analyzing Microarray Data with the mAdb System - Description and Sign up for a two half-day hands-on class using mAdb.
  - Statistical Analysis of Microarray Data - Description and Sign up for an overview of statistical issues and experimental design with a hands-on lab with BRB ArrayTools.

- mAdb Training Documentation
  - Introduction to mAdb (CIT class #411) Training Slides with Labs: PowerPoint or PDF
    Updated Tuesday, 28-Sep-2004 17:23:08 EDT
  - Analyzing Microarray Data with the mAdb System (CIT class #412) Training Slides
    - Lecture Slides: PowerPoint or PDF
      Updated Monday, 22-Nov-2004 18:18:45 EST
    - Hands-on Labs: PowerPoint or PDF
      Updated Tuesday, 16-Nov-2004 11:10:23 EST

- mAdb Reference Documentation
  - Increasing Upload Speed with Internet Explorer on the PC: Word or PDF
    Updated Wednesday, 14-May-2003 10:23:06 EDT
  - Uploading Affymetrix Data to mAdb: PDF
    Updated Thursday, 27-May-2004 16:04:17 EDT
    NOTE: You must request permission from mAdb support before uploading Affymetrix Data.

- Must request Affy privileges be turned on for your account

II. Putting your data in mAdb

Data Upload

- Login to mAdb Gateway page
  - change password if first-time user (case sensitive)
- Create project - logical organization for arrays
- Grant project access to others (if desired)
- Return to gateway and use Upload Array data link
- Select **type** of array for project
  - Spotted OR
  - Affymetrix (need to request permission via e-mail for first usage so we can give you needed parameters)

mAdb Gateway- link for User Profile Management

mAdb Gateway

NEW Create/Manage Projects link under Management Tools below. From there you can Create, Edit and Delete (empty projects) projects as well as Manage Access to projects.

Choose one or more Projects, select a Tool and Continue
or access previously extracted data located in **ncidemo**'s:
Temporary or Permanent area

Projects:

- XX guest - Time Course Demo Set #1
- XX guest - Time Course Demo Set #2
- XX guest - Repeats and Reciprocal Retests Demo Set #3
- XX guest - Multiple Types Demo Set #4
- AU ncidemo - my project
- AU ncidemo - Oligo and cDNA

Note: Tools marked with "*" only support selection of one project

Tool: Project Summaries Report

Continue

Uploading Links

- Upload Array data
- Status of Uploads
- Upload Identifier lists
- Manage Identifier lists

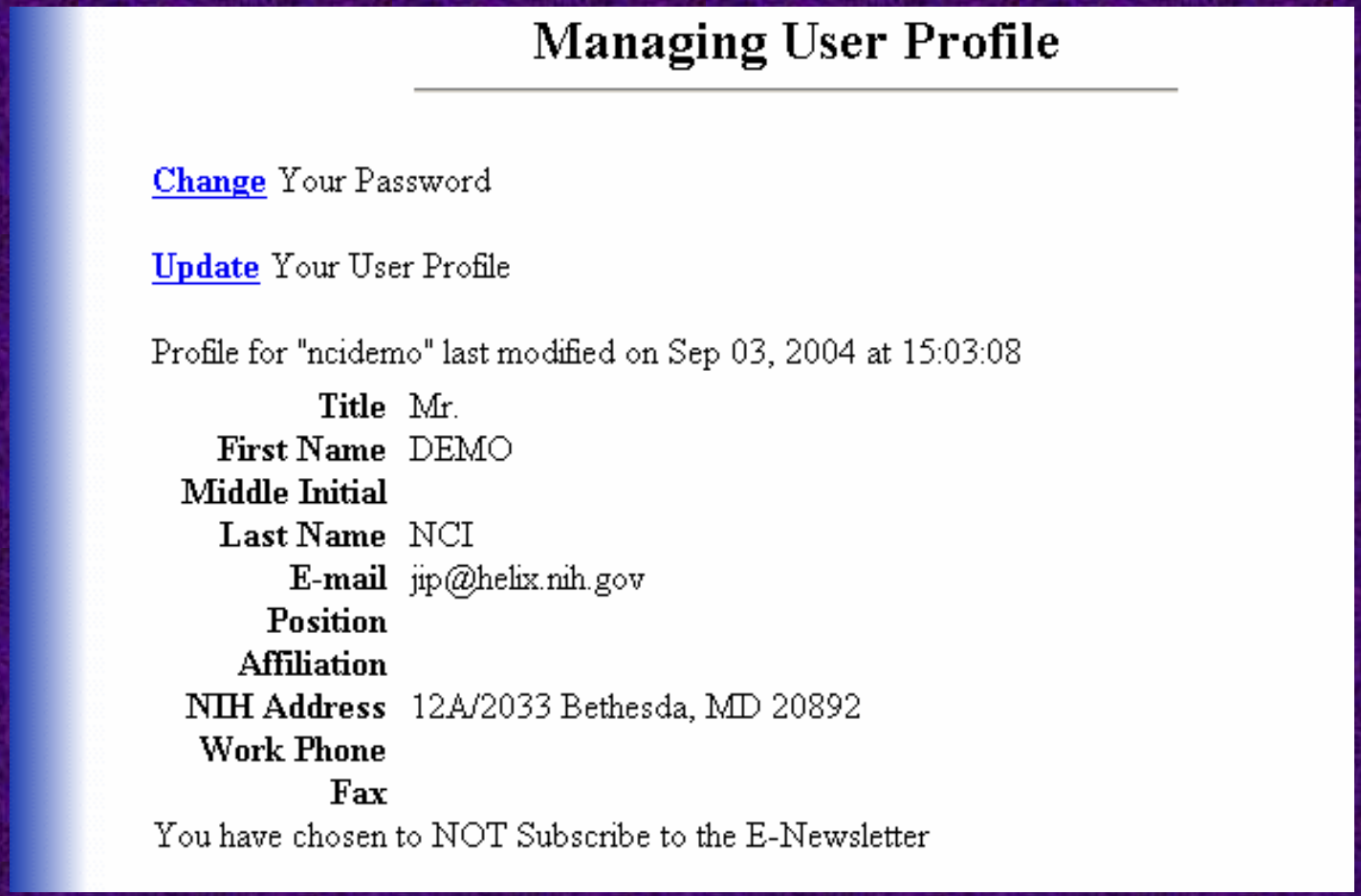
Management Tools

- Create/Manage Projects
- Manage User Profile

Access Training/Public Datasets

Access Additional Public Datasets

User Profile Management



Managing User Profile

Change Your Password

Update Your User Profile

Profile for "ncidemo" last modified on Sep 03, 2004 at 15:03:08

Title	Mr.
First Name	DEMO
Middle Initial	
Last Name	NCI
E-mail	jip@helix.nih.gov
Position	
Affiliation	
NIH Address	12A/2033 Bethesda, MD 20892
Work Phone	
Fax	

You have chosen to NOT Subscribe to the E-Newsletter

mAdb Gateway- link for Project Creation & Management

mAdb Gateway

NEW Create/Manage Projects link under Management Tools below. From there you can Create, Edit and Delete (empty projects) projects as well as Manage Access to projects.

Choose one or more Projects, select a Tool and Continue
or access previously extracted data located in **ncidemo**'s:
Temporary or Permanent area

Projects:

- XX guest - Time Course Demo Set #1
- XX guest - Time Course Demo Set #2
- XX guest - Repeats and Reciprocal Retests Demo Set #3
- XX guest - Multiple Types Demo Set #4
- AU ncidemo - my project
- AU ncidemo - Oligo and cDNA

Note: Tools marked with "*" only support selection of one project

Tool: Project Summaries Report

Continue

Uploading Links

- Upload Array data
- Status of Uploads
- Upload Identifier lists
- Manage Identifier lists

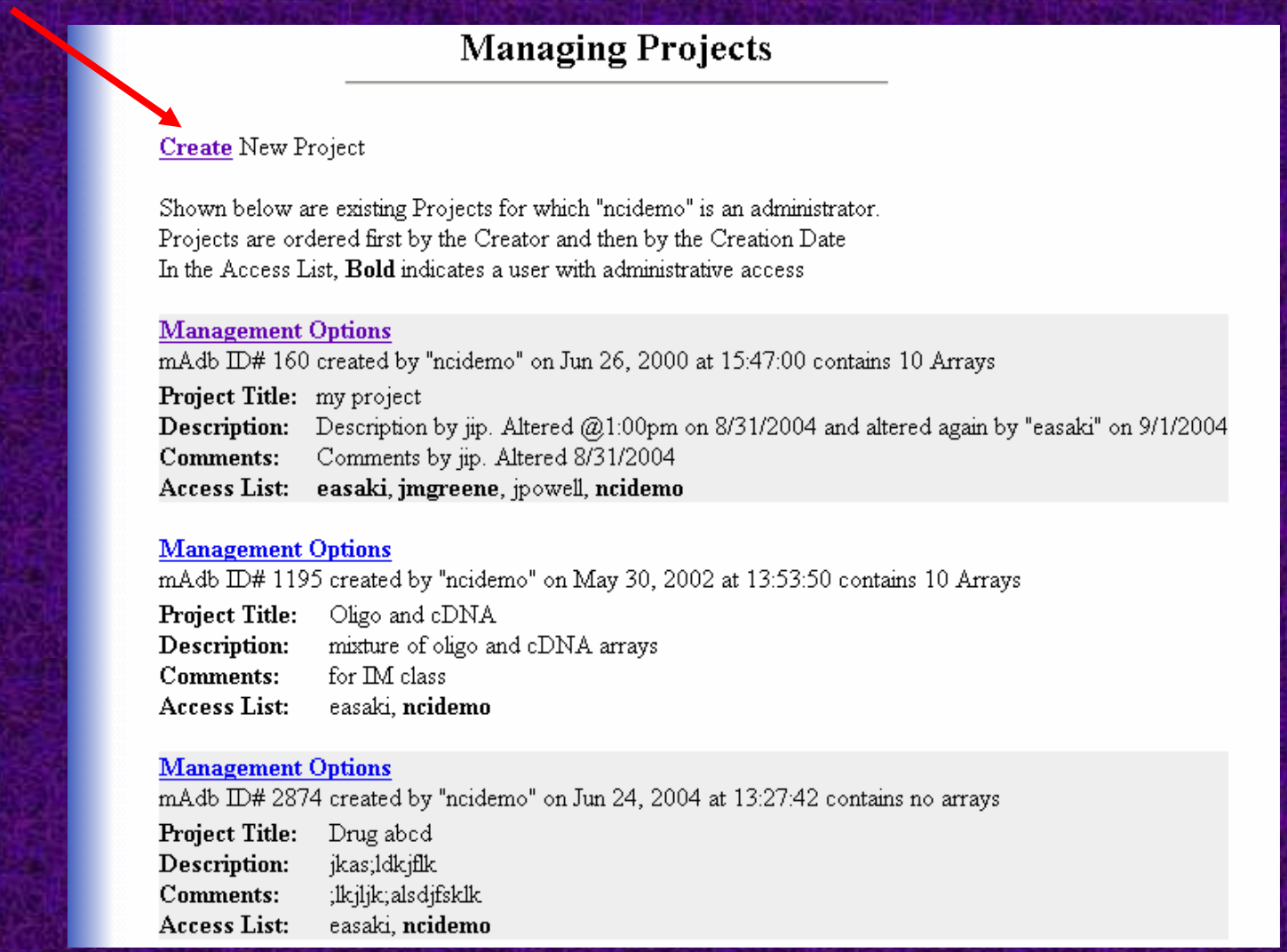
Management Tools

- Create/Manage Projects
- Manage User Profile

Access **Training/Public Datasets**

Access **Additional Public Datasets**

Managing Projects



Managing Projects

Create New Project

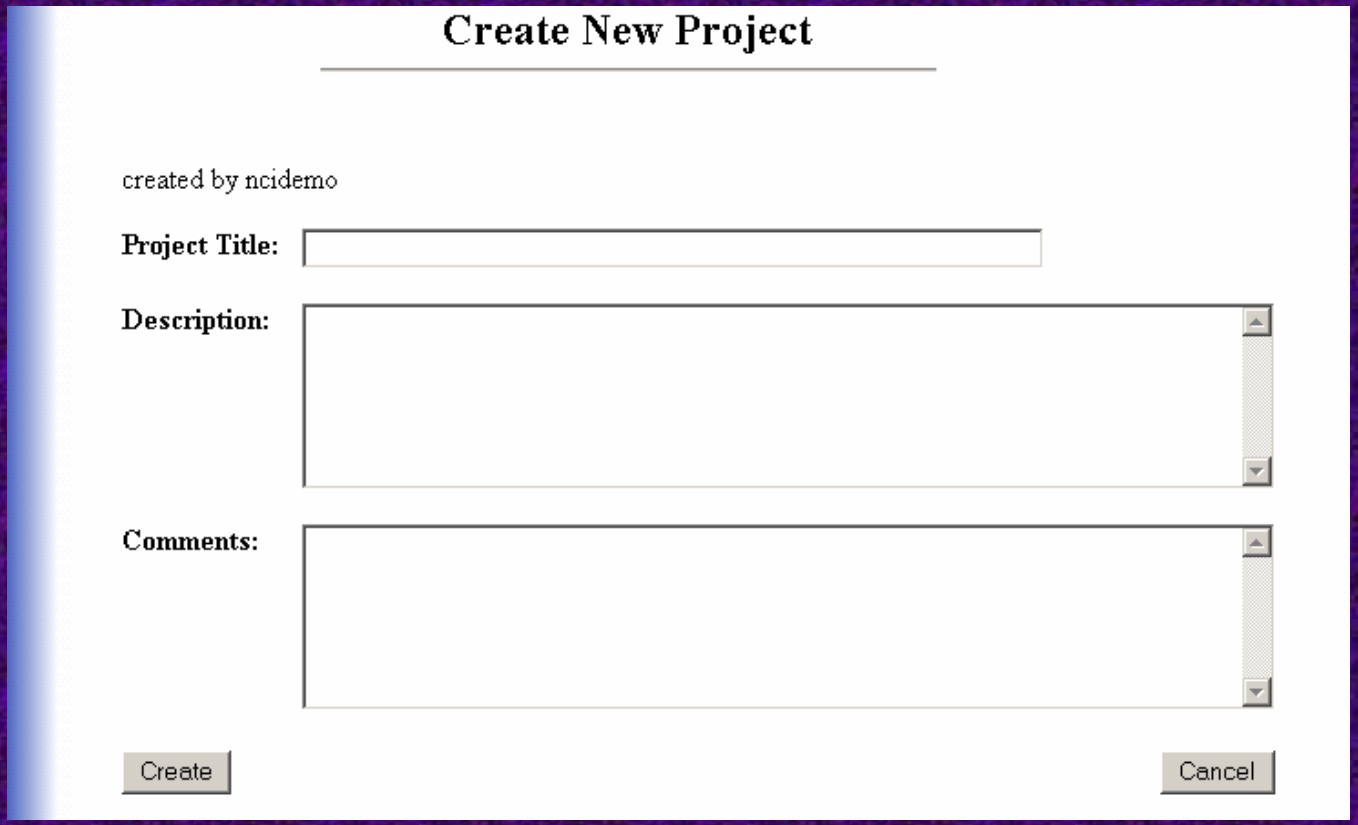
Shown below are existing Projects for which "ncidemo" is an administrator.
Projects are ordered first by the Creator and then by the Creation Date
In the Access List, **Bold** indicates a user with administrative access

Management Options
mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays
Project Title: my project
Description: Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004
Comments: Comments by jip. Altered 8/31/2004
Access List: **easaki**, **jmgreene**, jpowell, **ncidemo**

Management Options
mAdb ID# 1195 created by "ncidemo" on May 30, 2002 at 13:53:50 contains 10 Arrays
Project Title: Oligo and cDNA
Description: mixture of oligo and cDNA arrays
Comments: for IM class
Access List: easaki, **ncidemo**

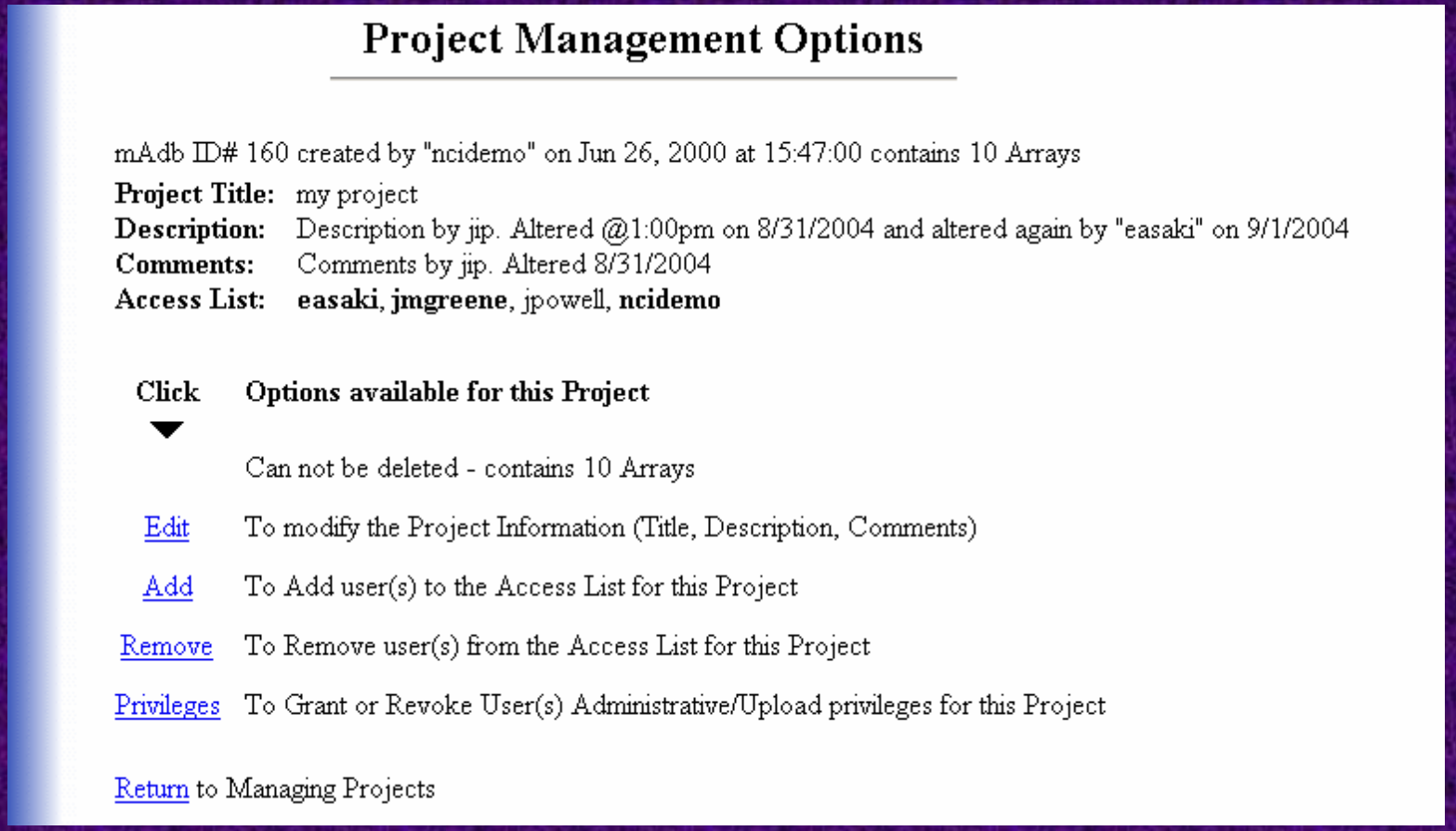
Management Options
mAdb ID# 2874 created by "ncidemo" on Jun 24, 2004 at 13:27:42 contains no arrays
Project Title: Drug abcd
Description: jkas;ldkjflk
Comments: ;lkjljk;alsdjfsklk
Access List: easaki, **ncidemo**

Create New Project



- **A Project is a logical grouping of your arrays**

Project Management Options



Project Management Options

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

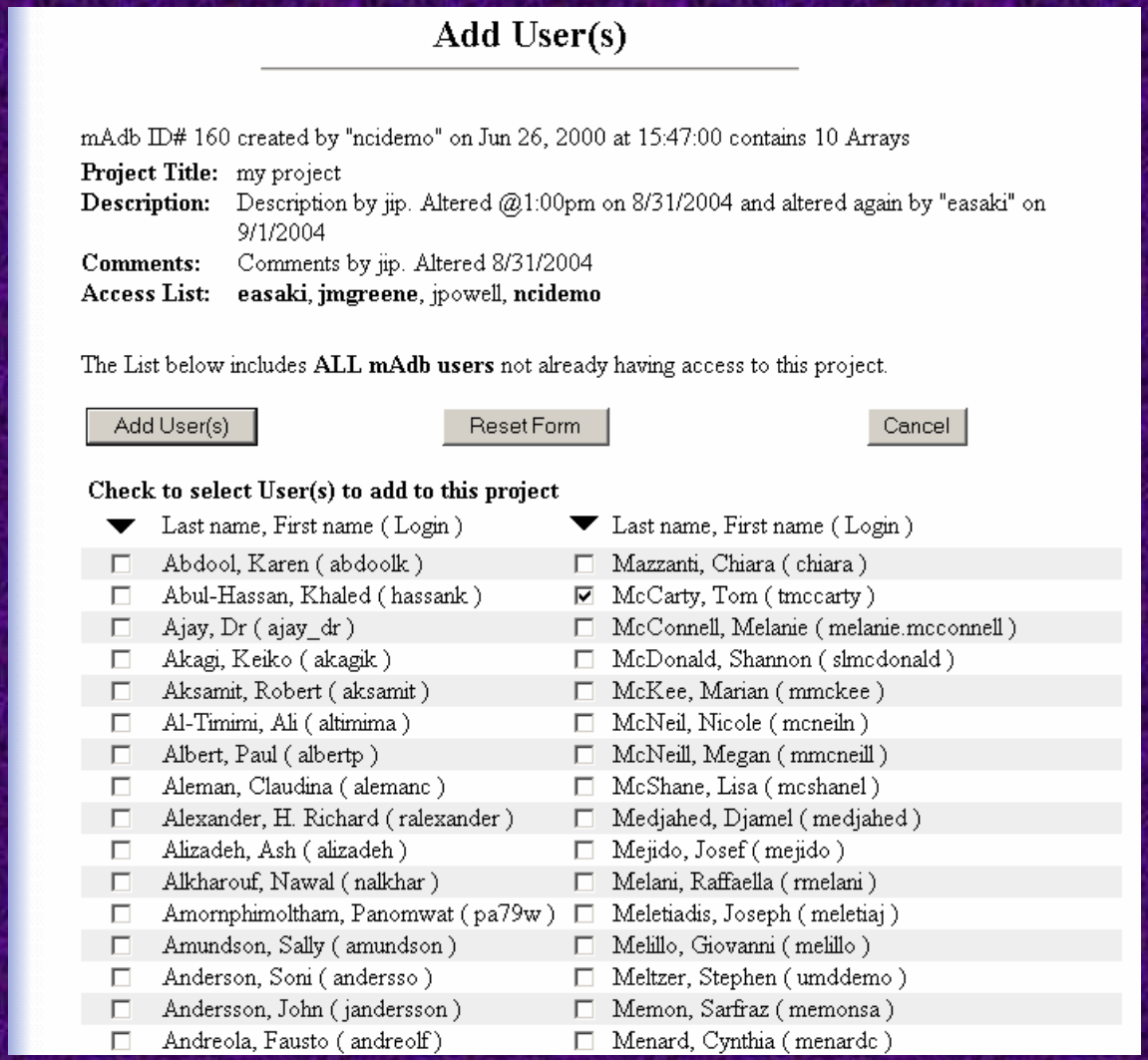
Project Title: my project
Description: Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004
Comments: Comments by jip. Altered 8/31/2004
Access List: **easaki**, **jmgreene**, jpowell, **ncidemo**

Click ▼	Options available for this Project
	Can not be deleted - contains 10 Arrays
Edit	To modify the Project Information (Title, Description, Comments)
Add	To Add user(s) to the Access List for this Project
Remove	To Remove user(s) from the Access List for this Project
Privileges	To Grant or Revoke User(s) Administrative/Upload privileges for this Project

Return to Managing Projects

Bold names on access list indicate administrative privileges for account

Project Access



Add User(s)

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

Project Title: my project
Description: Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004
Comments: Comments by jip. Altered 8/31/2004
Access List: **easaki**, **jmgreene**, jpowell, **ncidemo**

The List below includes **ALL mAdb users** not already having access to this project.

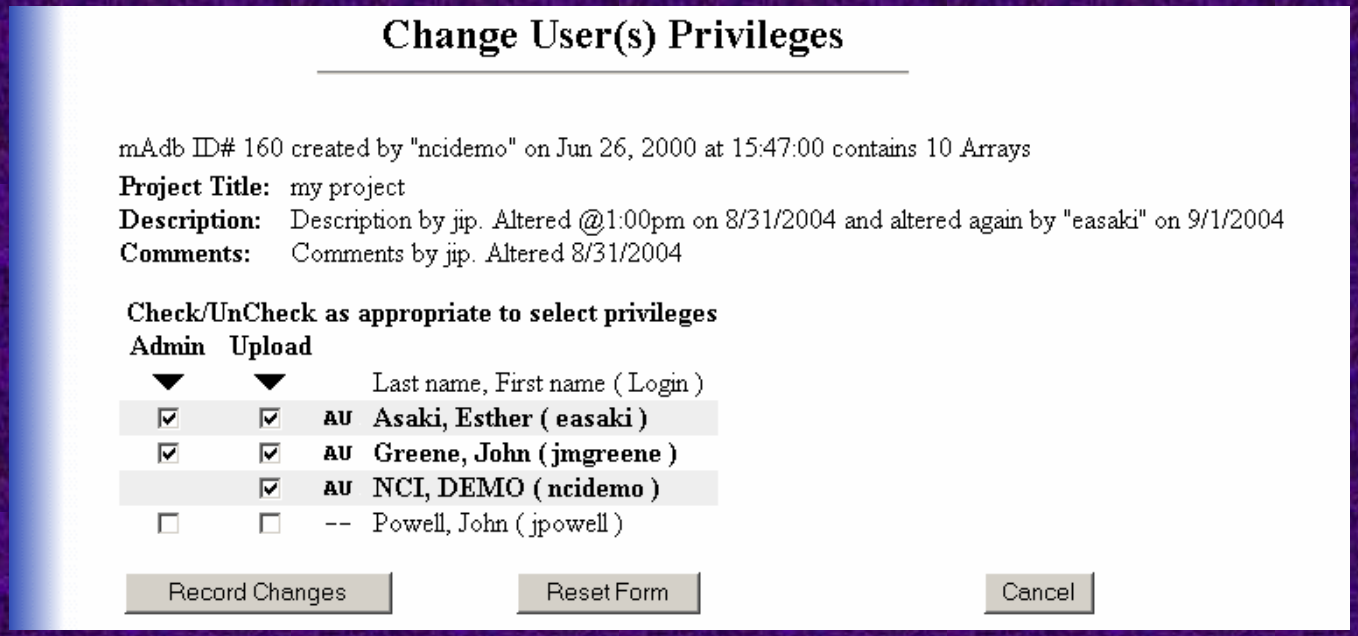
Add User(s) | Reset Form | Cancel

Check to select User(s) to add to this project

Last name, First name (Login)	Last name, First name (Login)
☐ Abdool, Karen (abdoolk)	☐ Mazzanti, Chiara (chiara)
☐ Abul-Hassan, Khaled (hassank)	☑ McCarty, Tom (tmccarty)
☐ Ajay, Dr (ajay_dr)	☐ McConnell, Melanie (melanie.mcconnell)
☐ Akagi, Keiko (akagik)	☐ McDonald, Shannon (slmcdonald)
☐ Aksamit, Robert (aksamit)	☐ McKee, Marian (mmckee)
☐ Al-Timimi, Ali (altimima)	☐ McNeil, Nicole (mcneiln)
☐ Albert, Paul (albertp)	☐ McNeill, Megan (mmcneill)
☐ Aleman, Claudina (alemanc)	☐ McShane, Lisa (mcshanel)
☐ Alexander, H. Richard (ralexander)	☐ Medjahed, Djamel (medjahed)
☐ Alizadeh, Ash (alizadeh)	☐ Mejido, Josef (mejido)
☐ Alkharouf, Nawal (nalkhar)	☐ Melani, Raffaella (rmelani)
☐ Amornphimoltham, Panomwat (pa79w)	☐ Meletiadis, Joseph (meletiaj)
☐ Amundson, Sally (amundson)	☐ Melillo, Giovanni (melillo)
☐ Anderson, Soni (andersso)	☐ Meltzer, Stephen (umddemo)
☐ Andersson, John (jandersson)	☐ Memon, Sarfraz (memonsa)
☐ Andreola, Fausto (andreolf)	☐ Menard, Cynthia (menardc)

Adding a user allows that mAdb account holder to view your arrays in a project and work with the data to create filtered datasets

User Access Levels



Change User(s) Privileges

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

Project Title: my project
Description: Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004
Comments: Comments by jip. Altered 8/31/2004

Check/UnCheck as appropriate to select privileges

Admin	Upload		Last name, First name (Login)
☑	☑	AU	**Asaki, Esther (easaki)**
☑	☑	AU	**Greene, John (jmgreene)**
	☑	AU	**NCI, DEMO (ncidemo)**
☐	☐	--	Powell, John (jpowell)

Record Changes | Reset Form | Cancel

Access levels allow user to:

- View data
- Upload Arrays
- Administer access to arrays and edit project, array, and dataset descriptions

mAdb Tool Gateway- link for uploading

mAdb Gateway

NEW Create/Manage Projects link under Management Tools below. From there you can Create, Edit and Delete (empty projects) projects as well as Manage Access to projects.

Choose one or more Projects, select a Tool and Continue
or access previously extracted data located in **ncidemo**'s:
Temporary or Permanent area

Projects:

- XX guest - Time Course Demo Set #1
- XX guest - Time Course Demo Set #2
- XX guest - Repeats and Reciprocal Retests Demo Set #3
- XX guest - Multiple Types Demo Set #4
- AU ncidemo - my project
- AU ncidemo - Oligo and cDNA

Note: Tools marked with "*" only support selection of one project

Tool: Project Summaries Report

Continue

Uploading Links

- Upload Array data
- Status of Uploads
- Upload Identifier lists
- Manage Identifier lists

Management Tools

- Create/Manage Projects
- Manage User Profile

Access Training/Public Datasets

Access Additional Public Datasets

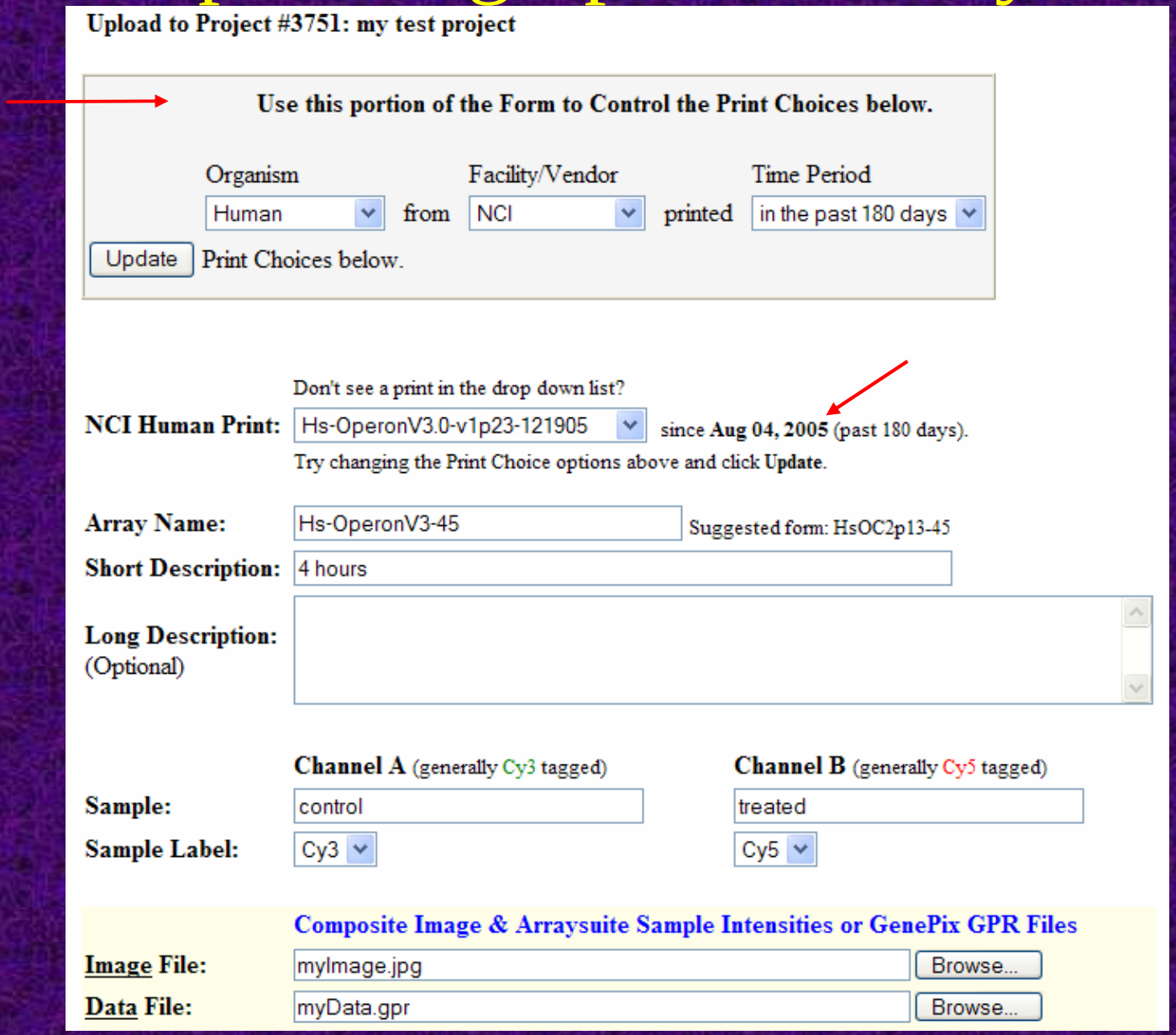
Spotted Array Data Upload

- Fill in experimental info for each array
  - Pick Print Set
  - Select image file of array
  - Select data file for array
- Submit and confirm upload
- Check upload status page to display progress
- Close browser when finished (for security)

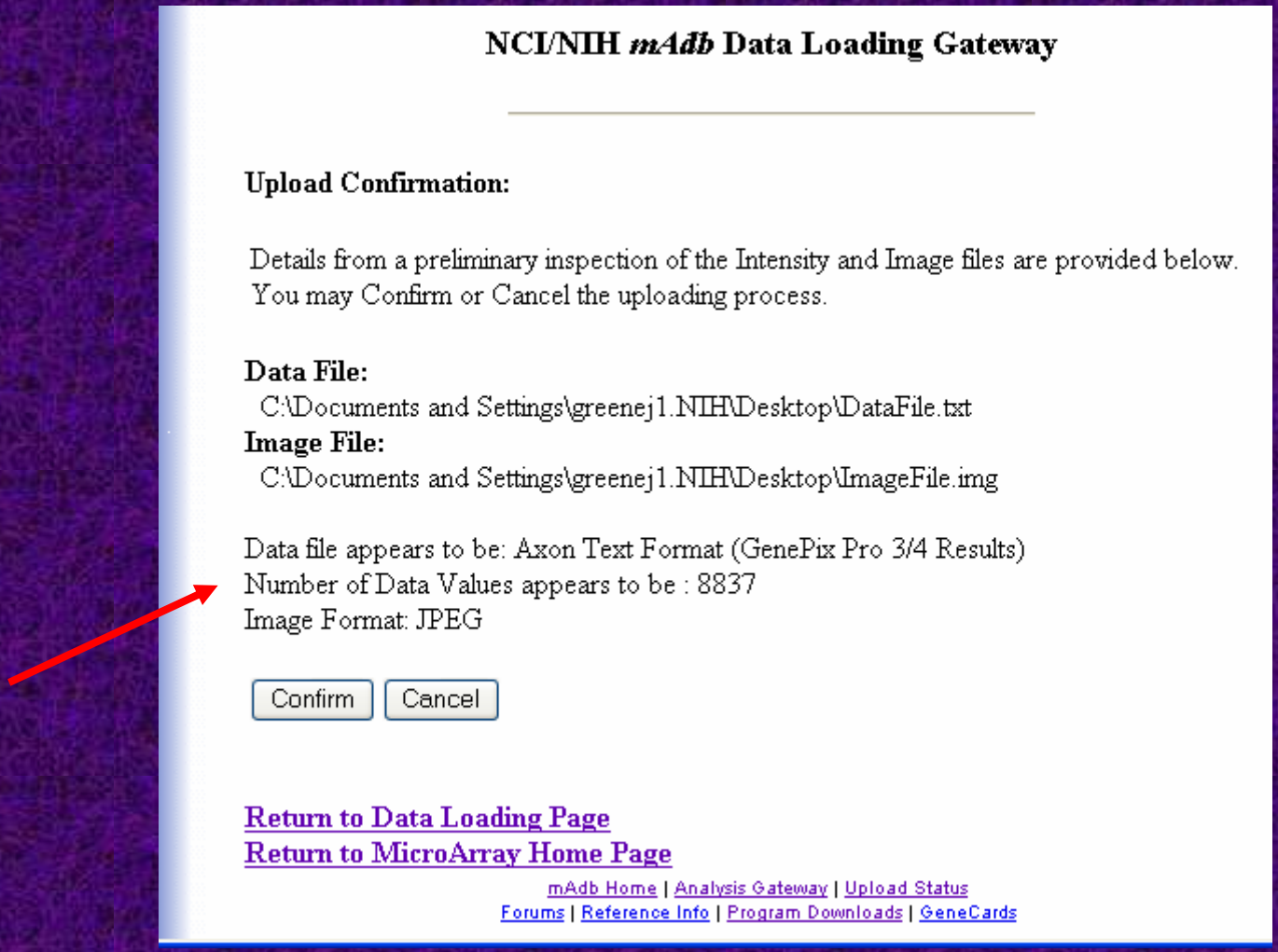
Affymetrix Data Upload

- Select:
  - Data File (Metrics - .txt file)
  - CEL file
- Fill in Experiment data
- Submit and confirm upload
- Check upload status page to display progress
- Close browser when finished (for security)

Uploading Spotted Arrays

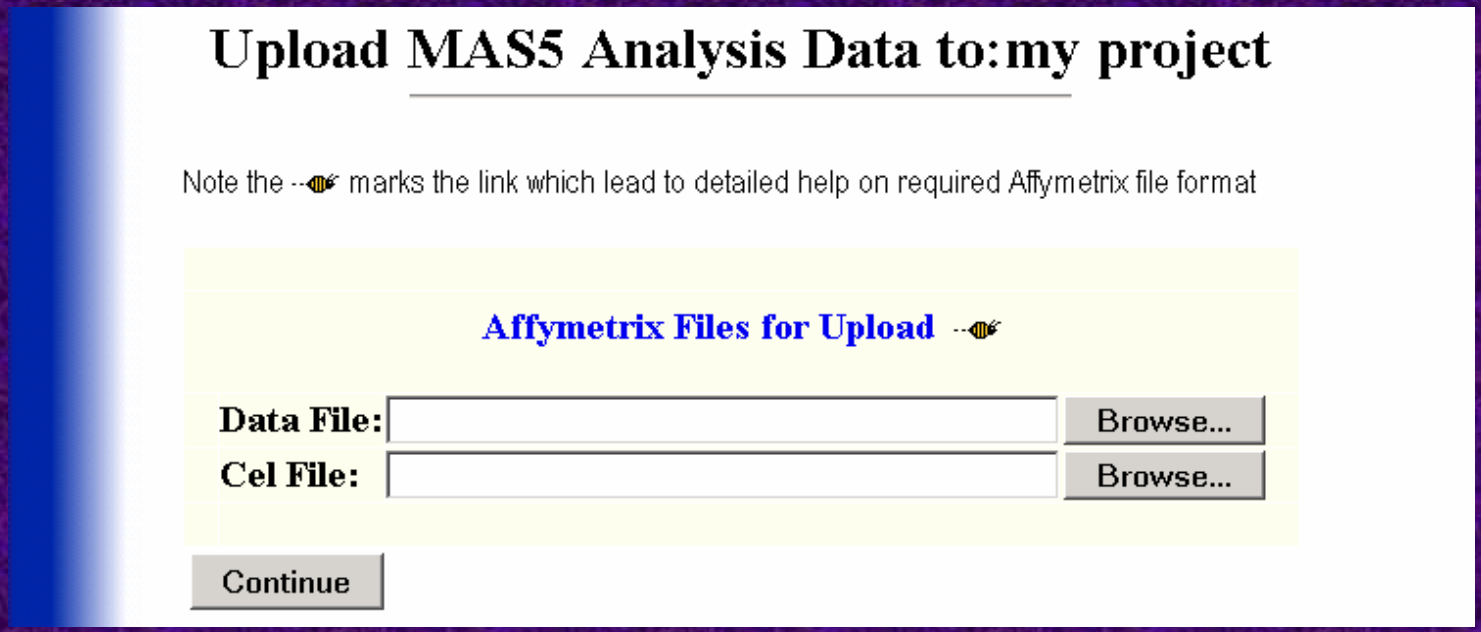


Confirming Upload



You should check that the image and file type appear correct and that the file line count is roughly equal to the number of spots on the array

Adding Affy Arrays



- Browse to Metrics (*.txt) file for the Data File box
- Browse to the corresponding .CEL file in second box

Adding Affy Arrays

Confirm Affymetrix Genechip Data

Experiment Information
You have uploaded Absolute Analysis data for a Human Genome Array U95A genechip.
The Data have not been scaled in your analysis.
Please check/complete the information on this page. Click the Confirm button to complete the upload process or use the Cancel button to abort and start again.

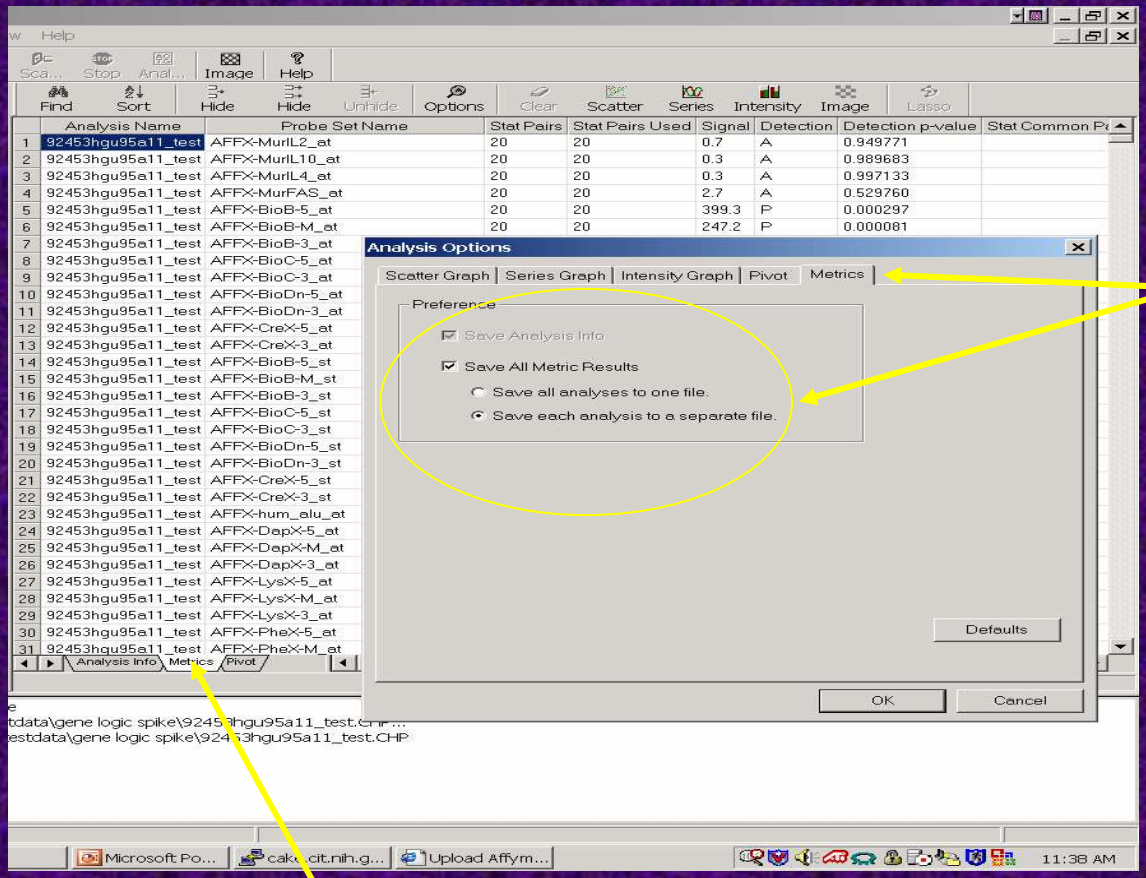
Uploaded Data File: C:\GeneChip\TESTDATA\Gene Logic Spike\92453hgu95a11_test.txt
Uploaded CEL File: C:\GeneChip\TESTDATA\Gene Logic Spike\92454hgu95a11.cel

Fields labled with ** are mandatoray.

Array Print Set:	U95A
Array Name: **	92453hgu95a11_test
Sample Type:	
Sample Description:	
Comments:	

Confirm Cancel

Affymetrix – CHP file

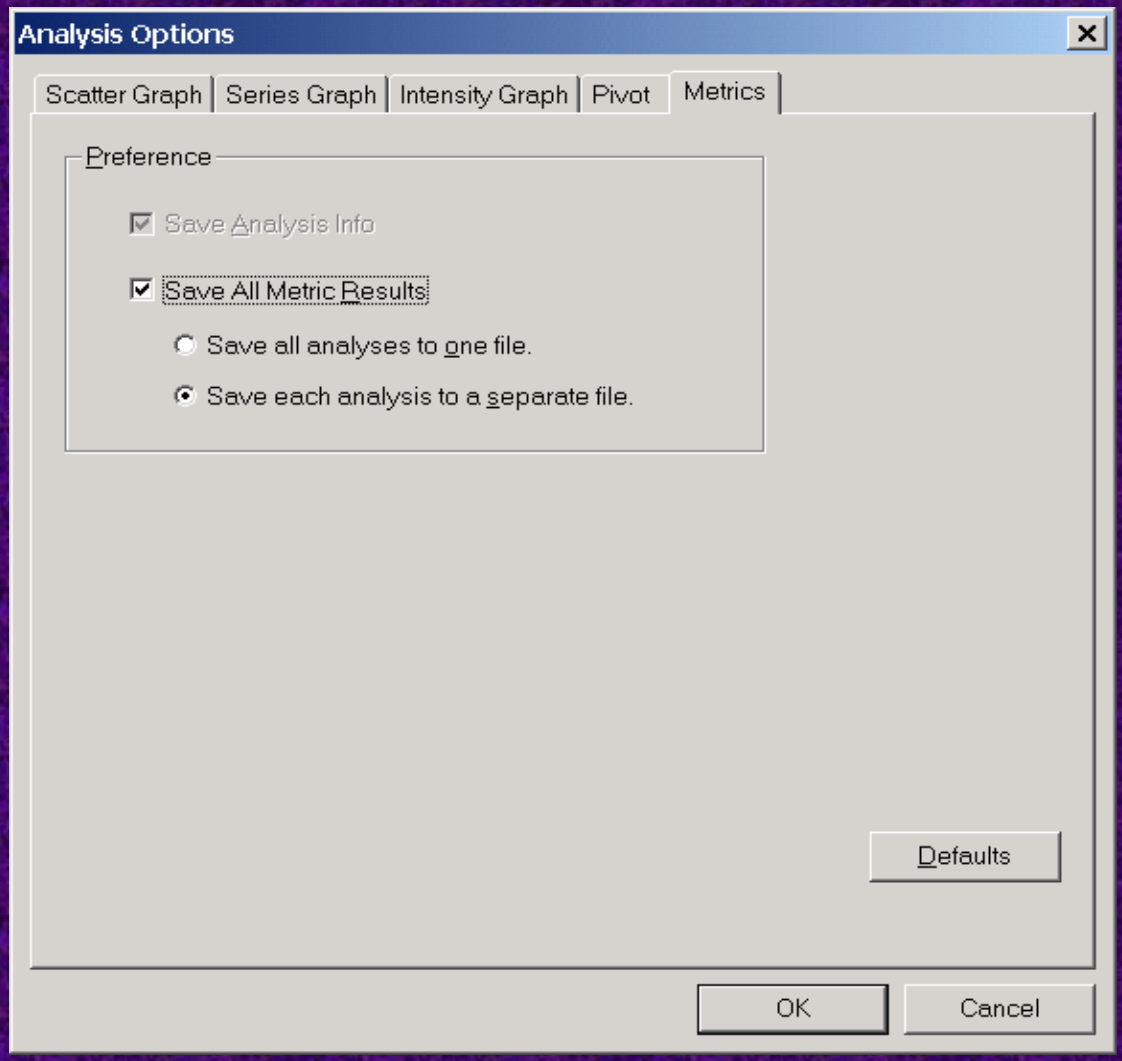


Set Metrics options:

- Save all Metric Results
- Save each analysis to a separate file

Select Metrics tab before saving

Affymetrix – CHP file Metrics options



Upload Status

- Shows your arrays and totals for all users
- Two step process:
  - Data is parsed and entered into Sybase db
  - Image is processed and stored
- You can work with data without waiting for image processing to finish

mAdb WEB Upload Status Report

Status Updated: Tue Sep 28 10:54:29 EDT 2004
(This page refreshes every 10 minutes)

Other mAdb WEB Upload Reports:
Graphical summary by month (past 12 months) or by day (past 90 days)
Details of arrays queued for processing
Details of arrays uploaded within the past 24 hours, 7 days, 30 days or all

mAdb login	Arrays	Status
ncidemo	0	Queued for/or loading into mAdb
Total all Users	0	Queued for/or loading into mAdb
ncidemo	0	Loaded; Queued for/or Image processing
Total all Users	0	Loaded; Queued for/or Image processing

Activity for the past 30 days

ncidemo	0	Processing completed
Total all Users	1037	Processing completed (332 Affymetrix, 705 Spotted)
ncidemo	0	Canceled, UnConfirmed, Bad Files/Rejected Submissions;
Total all Users	44	Canceled, UnConfirmed, Bad Files/Rejected Submissions

GenePix Analysis Notes

- Download correct GAL file from mAdb
- Carefully grid each block
- Allow program to “Find spots” and adjust spot size
- Set option to “Analyze absent spots”
- Adjust JPEG for desired contrast/brightness
- Analyze spots

Spotted Array Uploading Notes

- Include the slide number scratched on the slide as part of the Array Name, which will act as a unique identifier
- If you don't see your print in the drop down list, then adjust the search parameters and press "Show" button

Common Spotted Array Errors

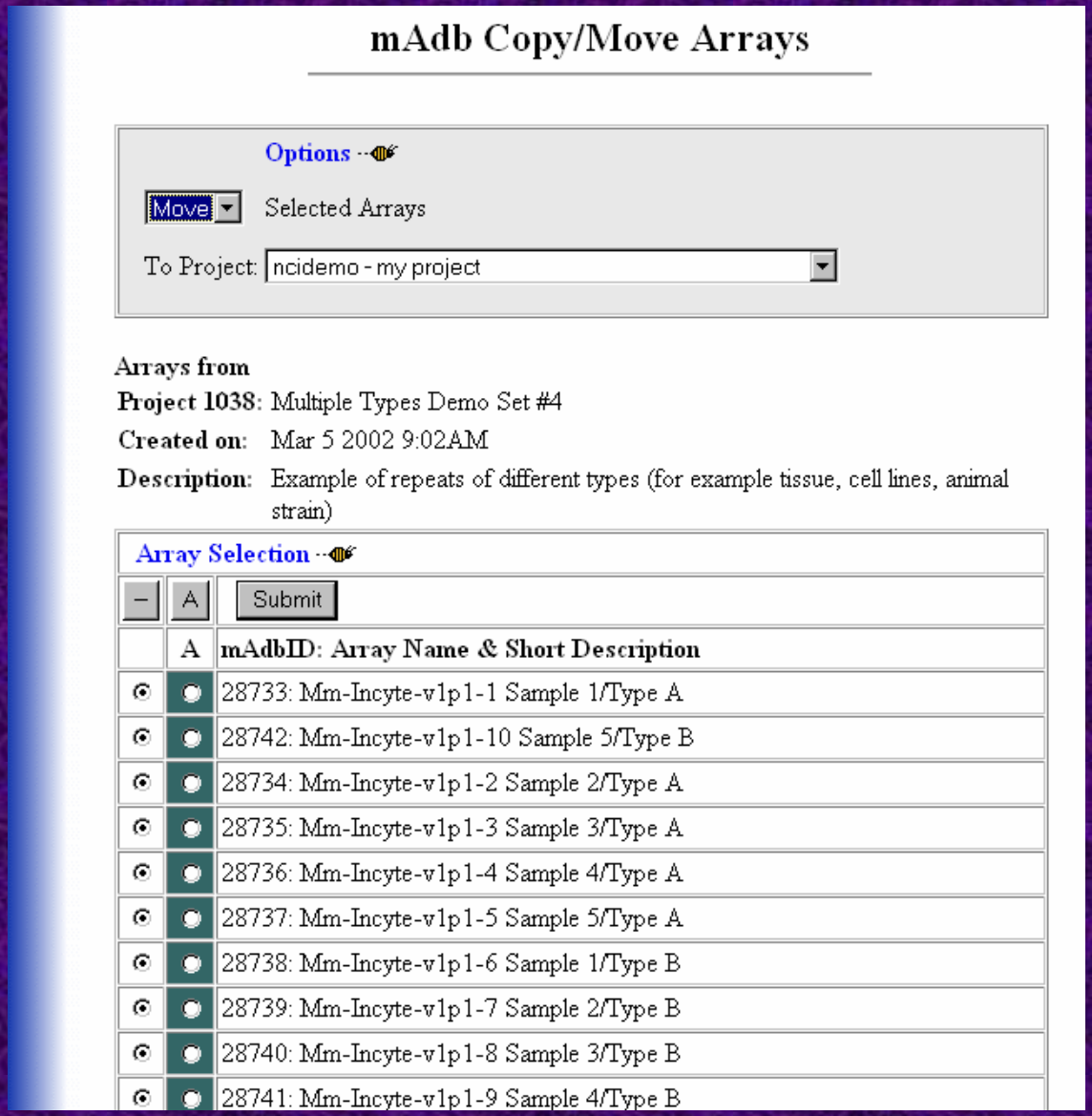
- Common Upload Errors
  - Choosing wrong print set
  - Loading GAL file, Excel file, or Set Up file in place of GenePix data (.gpr) file
  - Loading multi-image TIFF file instead of composite, single image JPEG or PICT file
- Common GenePix Errors
  - Setting incorrect option for “Analyze Absent Feature” (box should be checked) – results in truncated blocks
  - Deleting blocks
  - Gridding improperly

Affymetrix Analysis Notes

- Run chip through fluidics station to get CEL file
- Analyze CEL file (usually scale all spots to 500)
- With CHP file open, set analysis options on metrics tab as:
    - "Save All Metric Results"
    - "Save each analysis to a separate file"
- Click on Metric tab
- Save file as …. Xxxx.txt
- Note: If uploading comparison data, then upload absolute baseline data first.

Copy or move arrays between projects

- Accessible from the Gateway Tool menu
- Need administrative access to both projects
- Create a “trash” project to “delete” unwanted arrays



mAdb Copy/Move Arrays

Options

Move Selected Arrays

To Project: ncidemo - my project

Arrays from

Project 1038: Multiple Types Demo Set #4

Created on: Mar 5 2002 9:02AM

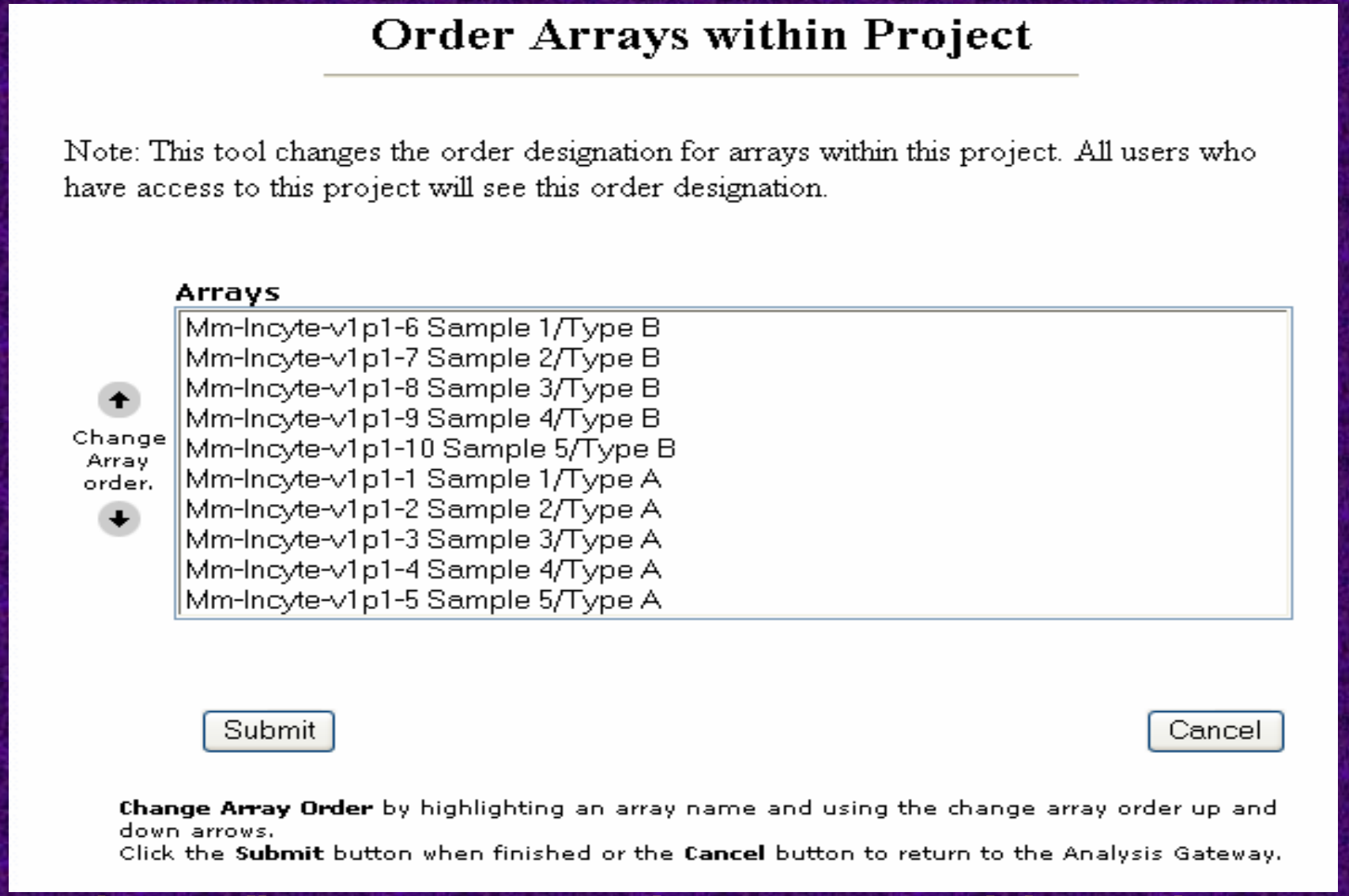
Description: Example of repeats of different types (for example tissue, cell lines, animal strain)

Array Selection

– A Submit

	A	mAdbID: Array Name & Short Description
◉	○	28733: Mm-Incyte-v1p1-1 Sample 1/Type A
◉	○	28742: Mm-Incyte-v1p1-10 Sample 5/Type B
◉	○	28734: Mm-Incyte-v1p1-2 Sample 2/Type A
◉	○	28735: Mm-Incyte-v1p1-3 Sample 3/Type A
◉	○	28736: Mm-Incyte-v1p1-4 Sample 4/Type A
◉	○	28737: Mm-Incyte-v1p1-5 Sample 5/Type A
◉	○	28738: Mm-Incyte-v1p1-6 Sample 1/Type B
◉	○	28739: Mm-Incyte-v1p1-7 Sample 2/Type B
◉	○	28740: Mm-Incyte-v1p1-8 Sample 3/Type B
◉	○	28741: Mm-Incyte-v1p1-9 Sample 4/Type B

Re-order arrays within a project



From the mAdb Gateway page, select a project and the “Order Arrays Within a Project” Tool and hit “Continue”

III. Evaluating Array Quality

- Signal definition
- Normalization
- Use of log base 2
- Project Summary Report
- Comprehensive Graphical Quality Report

mAdb Definitions

- Signal - refers to background corrected values (i.e.Target Intensity - Background Intensity).
- Defaults:
  - MEAN Intensity – MEDIAN background (for GenePix)
  - MEAN Intensity – MEAN background (for ArraySuite)
- Normalization factor – initially calculated so that the median overall ratio (Cy5 Signal/ Cy3 Signal) is adjusted to 1.0 (linear space) or 0.0 (log base 2) for each array. Spots with an extremely low signal are excluded from this calculation.

Need for Normalization of Ratios

- Unequal incorporation of labels (green Cy3 incorporates better than red Cy5)
- Unequal amounts of samples
- Unequal PMT voltage settings
- Different backgrounds
- Total brightness may differ between chips

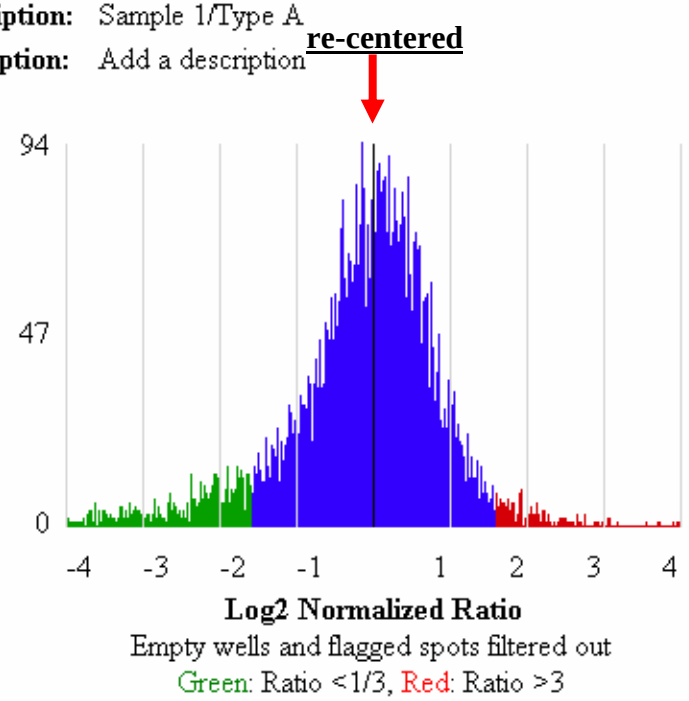
Why use ratios converted to log base 2?

- Makes variation of ratios more independent of absolute magnitude
- Symmetrical graphing – otherwise upregulated genes plotted from 1 to ∞ ; downregulated genes compressed between 0 and 1
- Clearer interpretation – negative numbers are downregulated genes; positive numbers are upregulated genes

mAdb Array Histogram

Comprehensive Graphical Report (Be Patient!)

Array: Mm-Incyte-v1p1-1
Short Description: Sample 1/Type A
Long Description: Add a description



Mean Signal		Median Bkg		Sgl/Bkg		Not	Normal.
Ch A	Ch B	Ch A	Ch B	Ch A	Ch B	Found	Factor**
326	455	110	84	3.0	5.4	30%	0.617

Normalization factor is calculated and multiplied against each ratio to re-center array distribution around 1 (linear), equal to 0 in log base 2

Project Summary

mAdb Project Summaries 1.0

Retrieve Array Summaries formatted for MS-Excel

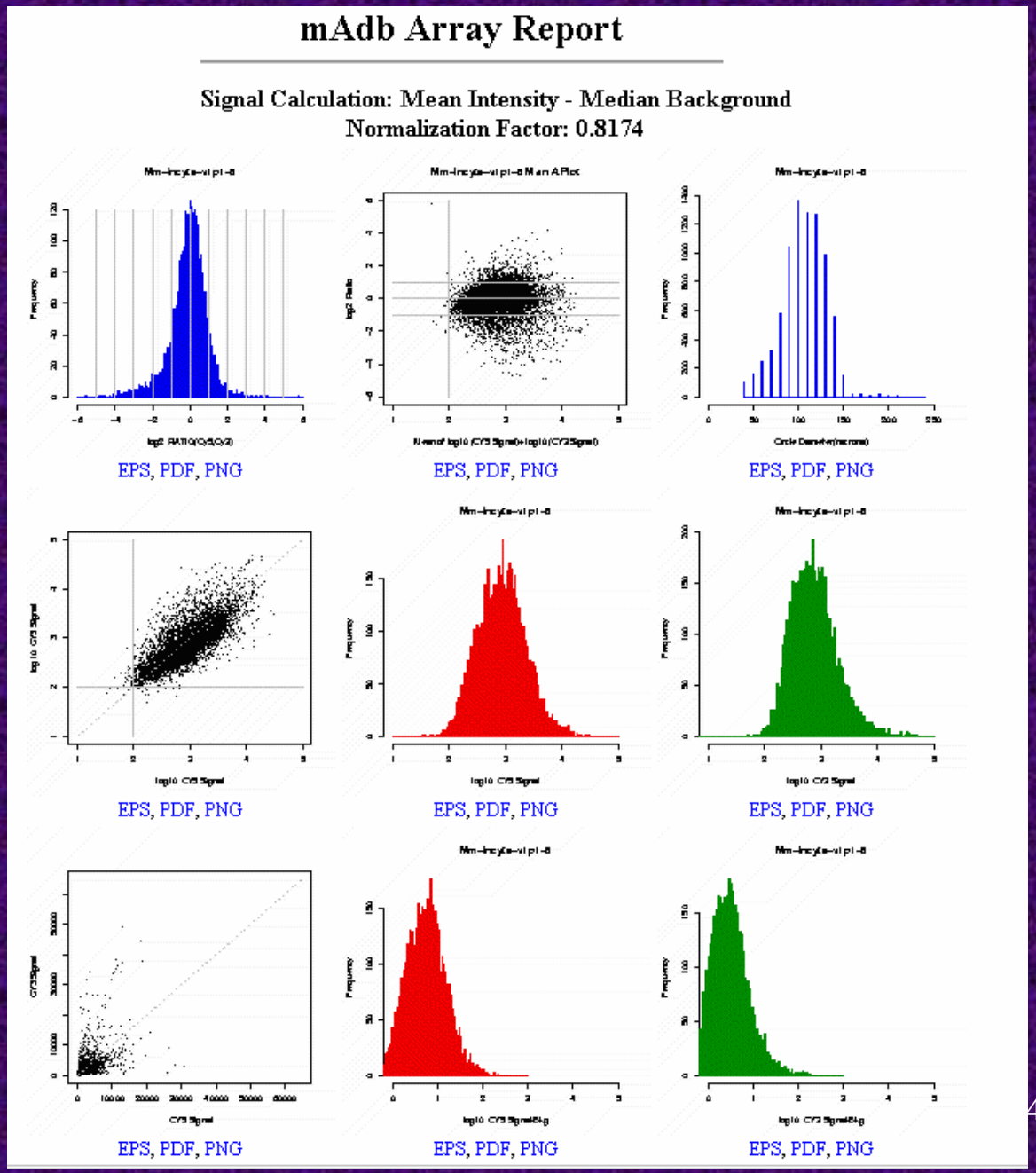
Edit **Project #1038**: Multiple Types Demo Set #4
Created on: Mar 05, 2002
Description: Example of repeats of different types (for example tissue, cell lines, animal strain)

		Summary Statistics								Array Information						
		Mean Signal Ch A	Mean Signal Ch B	Median Bkg Ch A	Median Bkg Ch B	Sgl/Bkg Ch A	Sgl/Bkg Ch B	% Found	Normal. Factor	mAdb ID	Uploaded	Array Print	Array	Probe A	Probe B	Short
Edit	1.	326	455	110	84	3.0	5.4	70%	0.626	28733	Mar 5 2002 9:07AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-1	Control	Sample 1	Sampl
Edit	2.	1677	2088	241	160	7.0	13.1	93%	0.769	28742	Mar 5 2002 9:24AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-10	Control	Sample 5/B	Sampl
Edit	3.	880	673	200	364	4.4	1.8	84%	1.055	28734	Mar 5 2002 9:10AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-2	Control	Sample 2	Sampl
Edit	4.	1056	1473	259	154	4.1	9.6	93%	0.658	28735	Mar 5 2002 9:11AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-3	Control	Sample 3	Sampl
Edit	5.	297	493	117	87	2.5	5.7	84%	0.542	28736	Mar 5 2002 9:13AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-4	Control	Sample 4	Sampl
Edit	6.	443	543	123	89	3.6	6.1	83%	0.708	28737	Mar 5 2002 9:15AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-5	Control	Sample 5	Sampl
Edit	7.	499	541	120	101	4.2	5.4	84%	0.858	28738	Mar 5 2002 9:17AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-6	Control	Sample 1/B	Sampl
Edit	8.	626	717	146	113	4.3	6.3	85%	0.890	28739	Mar 5 2002 9:21AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-7	Control	Sample 2/B	Sampl
Edit	9.	1280	1399	272	190	4.7	7.4	93%	0.830	28740	Mar 5 2002 9:22AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-8	Control	Sample 3/B	Sampl
Edit	10.	1113	1371	261	156	4.3	8.8	91%	0.779	28741	Mar 5 2002 9:23AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-9	Control	Sample 4/B	Sampl

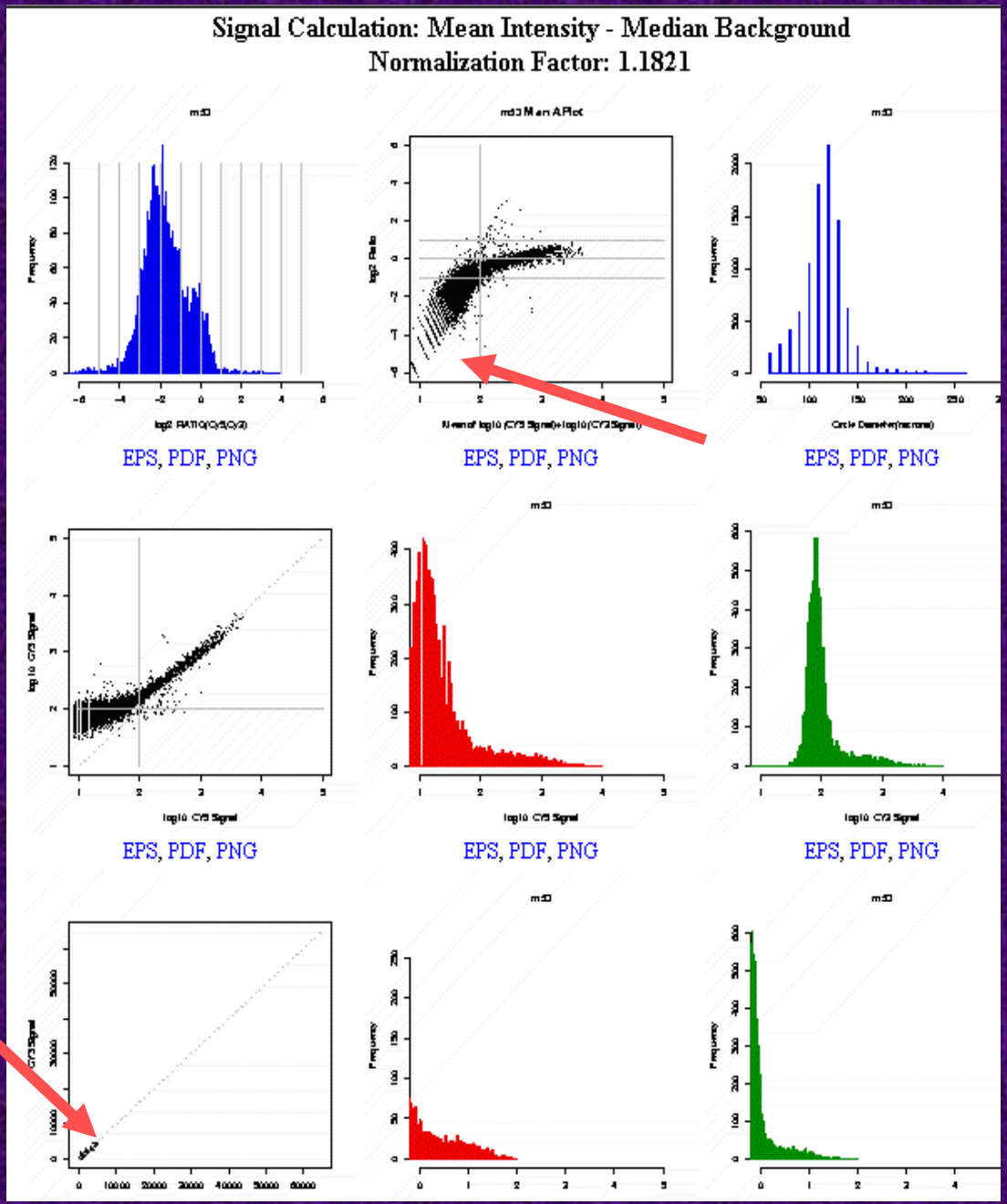
- Aid to QC – overall array statistics, links to histogram, array image
- If you have admin access to a project, can edit project and array descriptions from “Edit” links here

Comprehensive Graphical Quality Report

- Accessed from histogram display
- More QC parameters, including:
- M versus A plot
- spot size distribution
- log and linear plots of each channel
- signal intensity distribution
- signal/background distribution



Low Intensity/Channel Failure Example



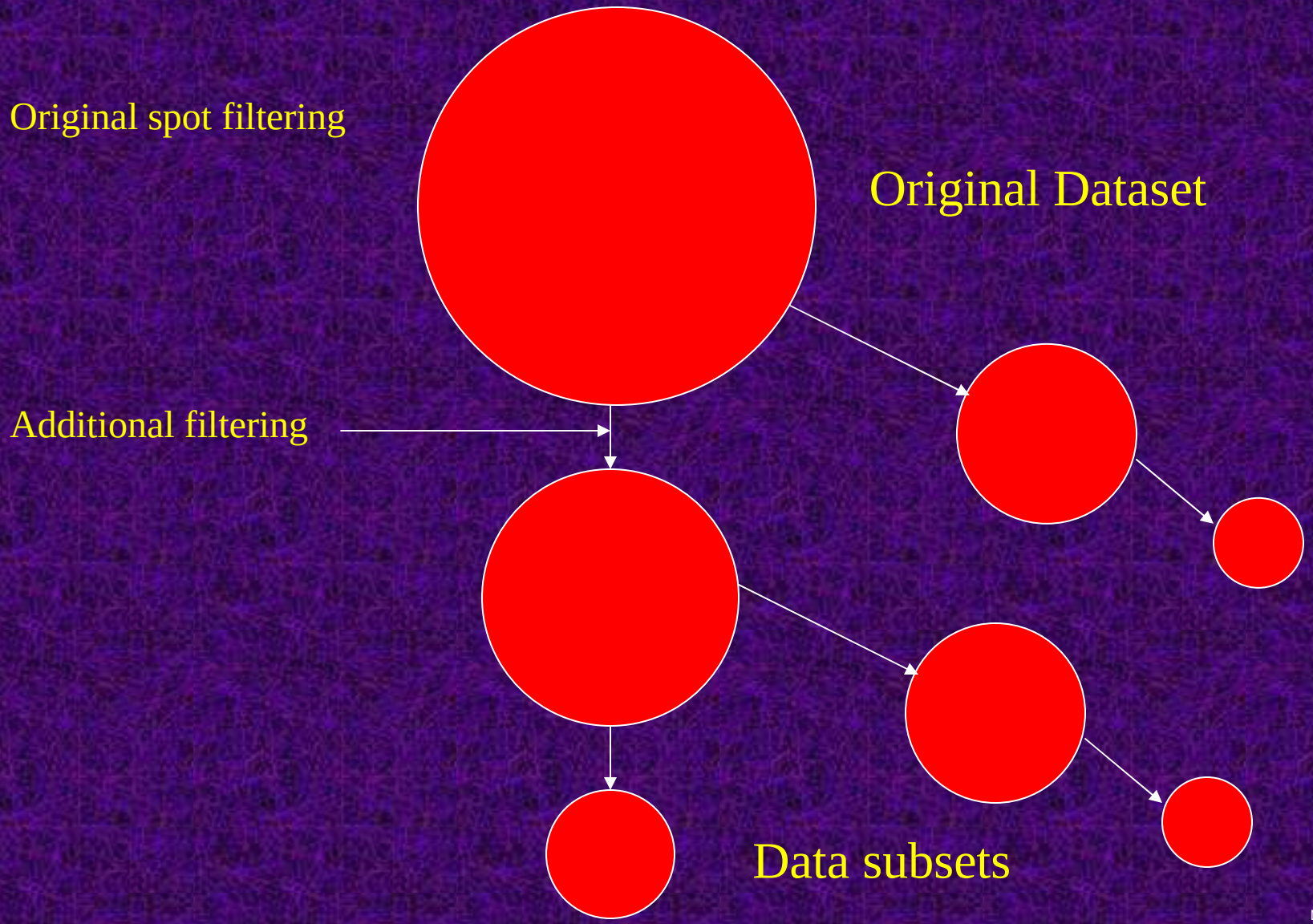
- M vs A plot – ratio distribution dependent upon signal strength; see a "tail" toward green spots
- Spot sizes small
- Overall signal strength very weak – not a good range of signals on Cy3/Cy5 linear plot
- Bulk of red signals less than 10
- FYI, max signal is 65,000

IV. Getting started with analysis

mAdb Analysis Paradigm:

1. Create project; Upload arrays to that project
2. Quality control – Project Summary and Graphical Reports
3. Create a filtered dataset:
   - Select arrays for analysis
   - Define quality parameters (minimum signal values, S/N, etc.)
   - Select normalization method, so different arrays can be compared
   - Align genes from different array layouts (based on well IDs)
4. Apply Data/Gene criteria filters, if desired, to create subset dataset(s)
5. Apply appropriate Analysis/Visualization Tools to the dataset(s)
6. Repeat Steps 3, 4, and 5 as desired
7. Interpret Datasets/Results

Dataset Structure -Filtering hierarchy /tree structure

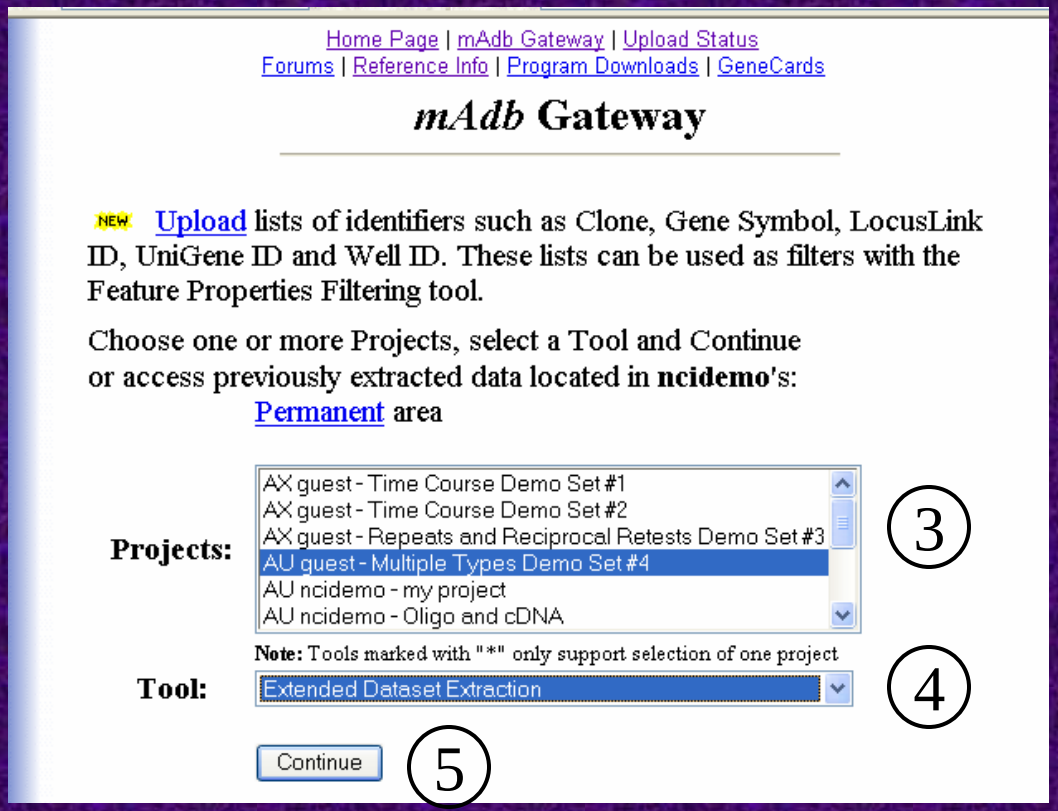


Lab 1 – Creating a filtered dataset

Goal: To start analyzing arrays using only high quality/reliable spots

Do NOT maximize the browser window, so multiple windows can be distinguished on the monitor.

Lab 1. Choosing Project and Extended Dataset Extraction Tool



1. Open a web browser and type the URL for the mAdb home page, **http://madb-training.cit.nih.gov** .

2. Click the first bullet on the home page, to access the **mAdb Gateway,** web page, shown at left. You will need to login the mAdb Gateway with the mAdb account as instructed.

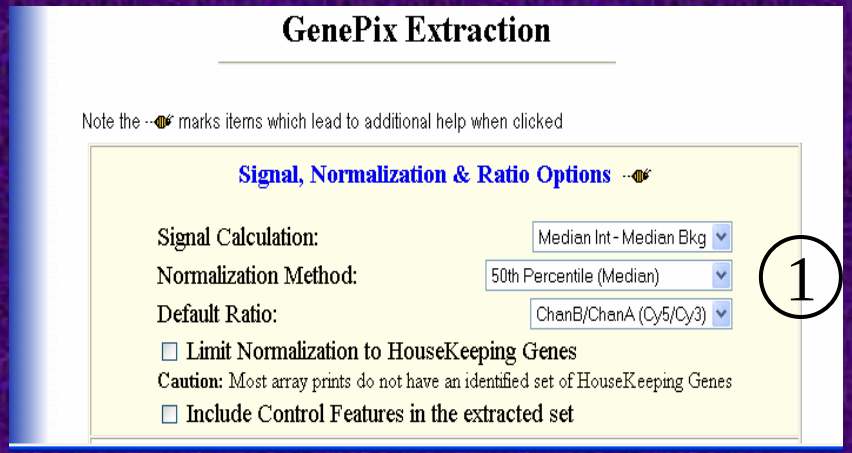
3. On the mAdb Gateway Web page, in the **Projects:** list, select the "**guest – Multiple Types Demo Set #4"** project

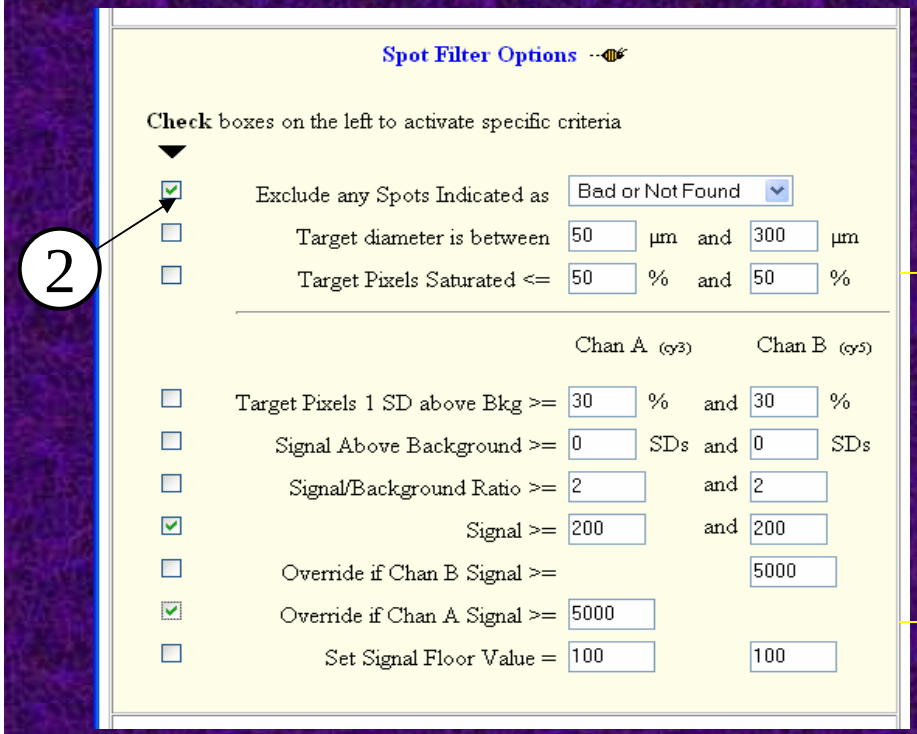
NOTE: You can select multiple projects by holding down the **Ctrl** key when you click on a project

4. On the **Tools:** menu just below, select "**Extended Dataset Extraction"**

5. Press the **Continue** button

Lab 1. Selecting Filtering Options



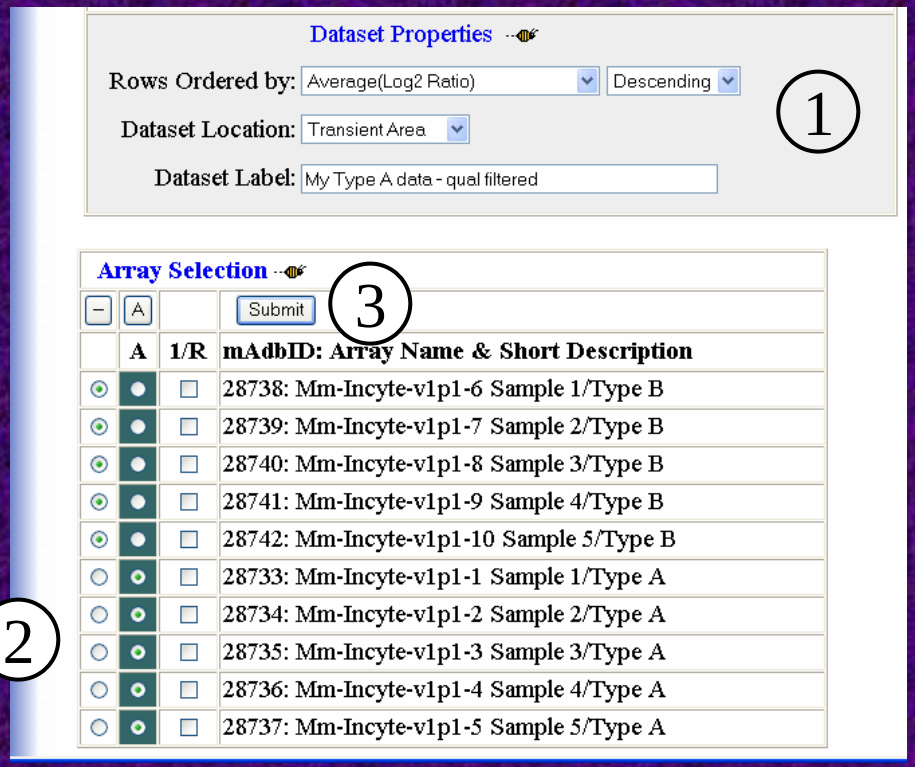


1. In the **Signal, Normalization, & Ratio Options** panel, choose **Signal Calculation:** Median Int – Median Bkg**, Normalization Method:** 50th Percentile (Median), and **Default Ratio:** ChanB/ChanA**.** Leave the checkboxes empty. Using this Normalization method, the output is re-normalized based on the spots which pass the filters.

2. In the **Spot Filter Options** panel, check the boxes on the left to activate the appropriate filter(s), and choose appropriate values by typing in numbers into the form elements to the right of each filter checkbox. For the purposes of this exercise, check:
   - Exclude any Spots indicated as **Bad or Not Found**
   - Signal >= **200** and **200**
   - Override if Chan B Signal >= **5000**
   - Override if Chan A Signal >= **5000**

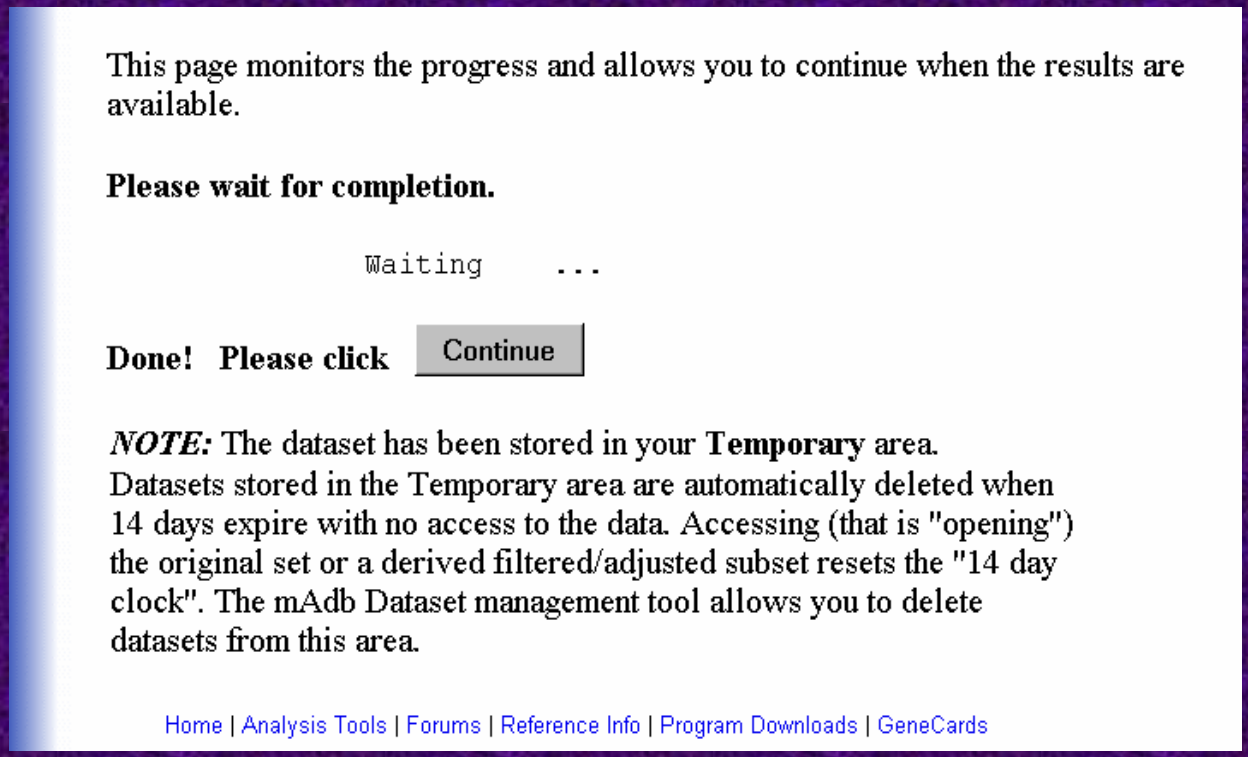
3. Go to next page of lab to choose arrays

Lab 1. Selecting Dataset Properties and Arrays



1. In the **Dataset Properties** panel, choose **Rows Ordered by: Average(Log2 Ratio)** and **Descending; Dataset Location: Transient Area,** and **Dataset Label:** "My Type A data – qual filtered" .
2. In the **Array Selection** panel, choose just the Type A arrays using the radio buttons under **A. N.B.** If a dye swap or reverse fluor, check the **1/R** box to take the reciprocal value of the ratio for direct comparison.
3. Press **Submit**

Lab 1. Waiting for Data Extraction …



Intermediate screen which monitors the data extraction process. When the creation of the working dataset is complete, the user can continue to the Data Display page.

Extended Tool: Signal, Normalization & Ratio Options:

- **Signal Calculation**

  Mean Intensity – Median Background

  Median Intensity – Median Background

- **Normalization**

  - None

  - 50th Percentile (Median)

  Applied to extracted spots (spots passing filter)

  All spots or only Housekeeping spots (on limited prints)

  - Pre-calculated 50th percentile (based on all spots)

  - Lowess non-linear normalization – in beta testing

- **Default Ratio**

  Chan B/Chan A (CY5/CY3),

  but for reverse fluor can choose Chan A/Chan B (CY3/CY5)

CIT

Spot Filter Options:

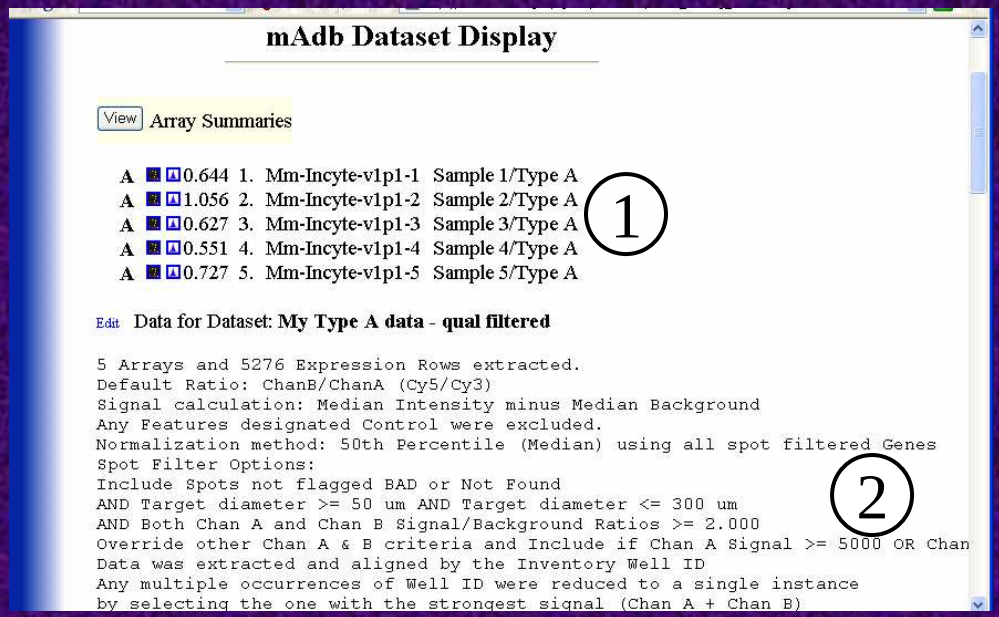
Important - Check box to Activate!

- **Exclude any Spots Flagged as *Bad Or Not Found, Bad***
- **Target diameter is between *xx and yy microns***
- **Target Pixels Saturated**
- **Target Pixels 1 Standard Deviation above background >= *N %***
- **Signal above background >= *N SDs (standard deviations)***
- **Signal/Background Ratio >= *N***
- **Signal >= N *(raw signals)***
- **Override bracketed criteria (in yellow above) if Chan B and /or A Signal >= *N***

Signal Floor

- When one channel has a very low signal and the other has a moderate or high signal, the resulting ratio value could be misleading (i.e. very high/low)
- To adjust such a highly skewed ratio, mAdb allows the user to set a floor (e.g. 100) for signals below a threshold

Lab 1. Main mAdb Dataset Display – Part 1

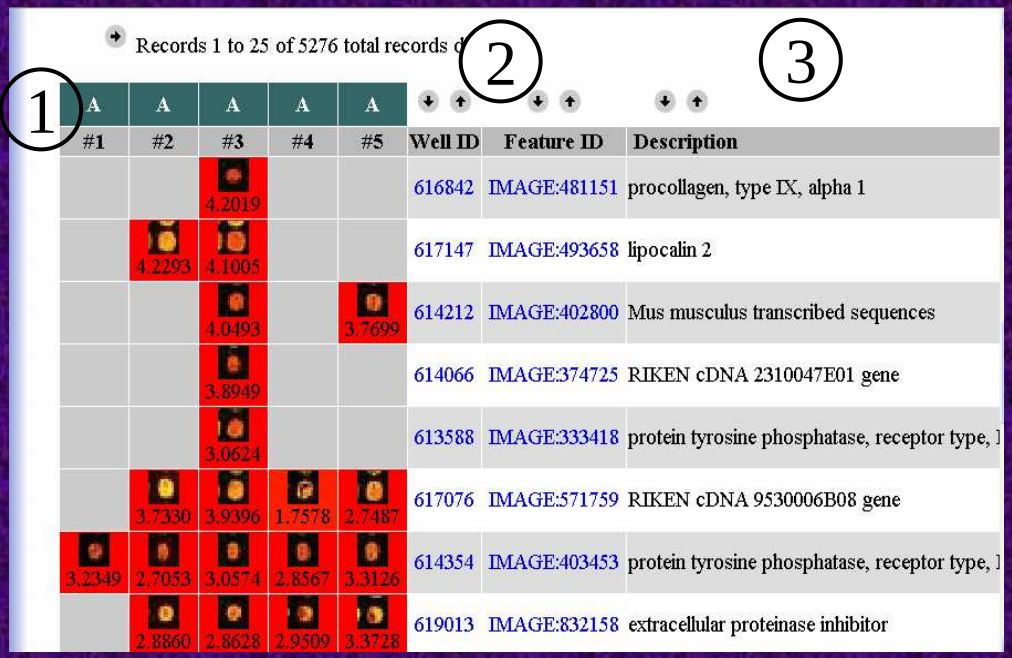


1. The listing at the top shows the array group, a link to the array image, a link to a histogram display, the re-calculated normalization factor (based on those spots which passed the quality filters), the array name, and the short description for all of the chosen arrays to be filtered

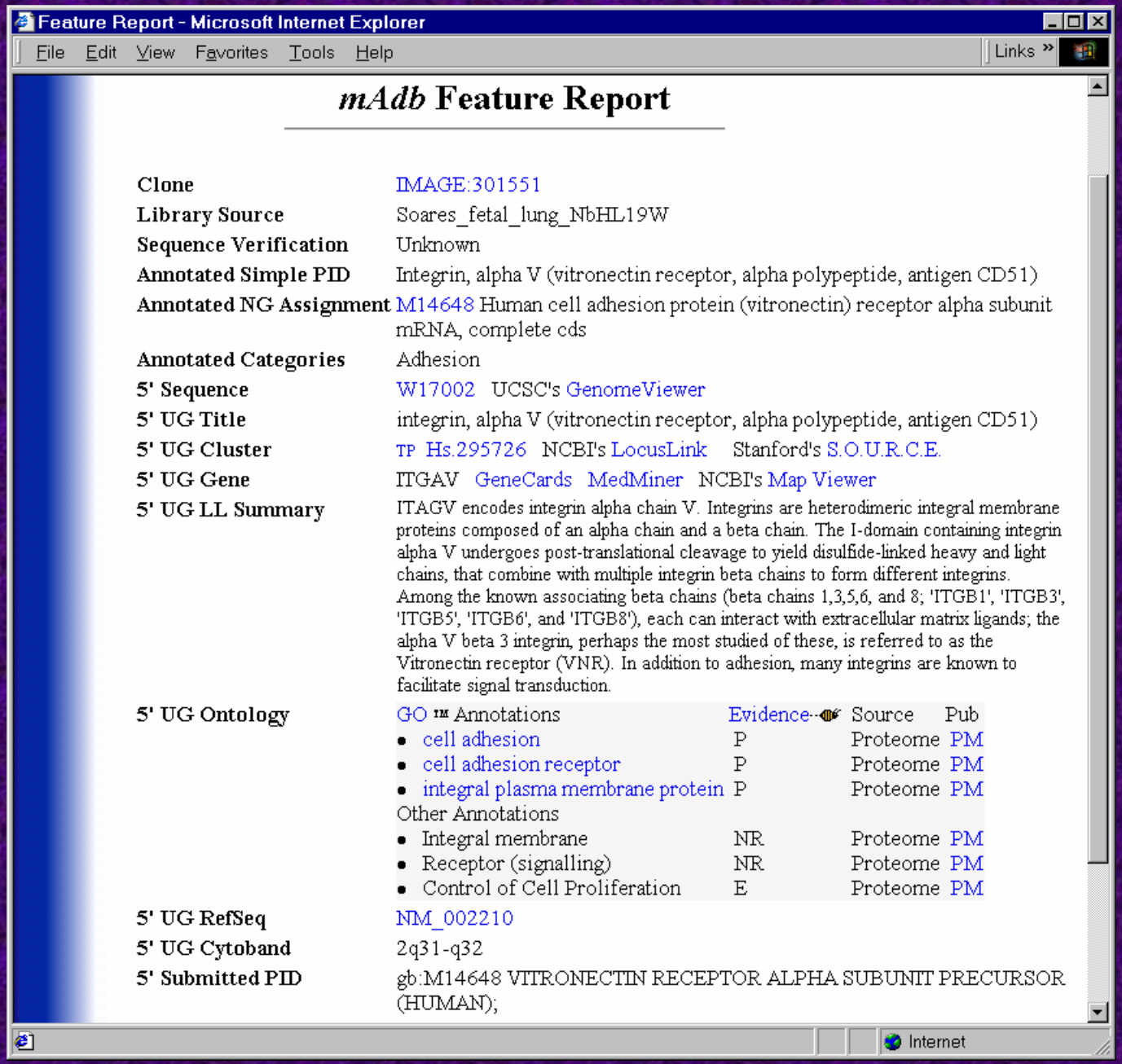
2. After the Dataset name (which can be **edit**ed with the link to the left), is the history of what was done in the preceding filtering step.

3. Go to the next page of the lab and scroll down to the bottom of the Web page.

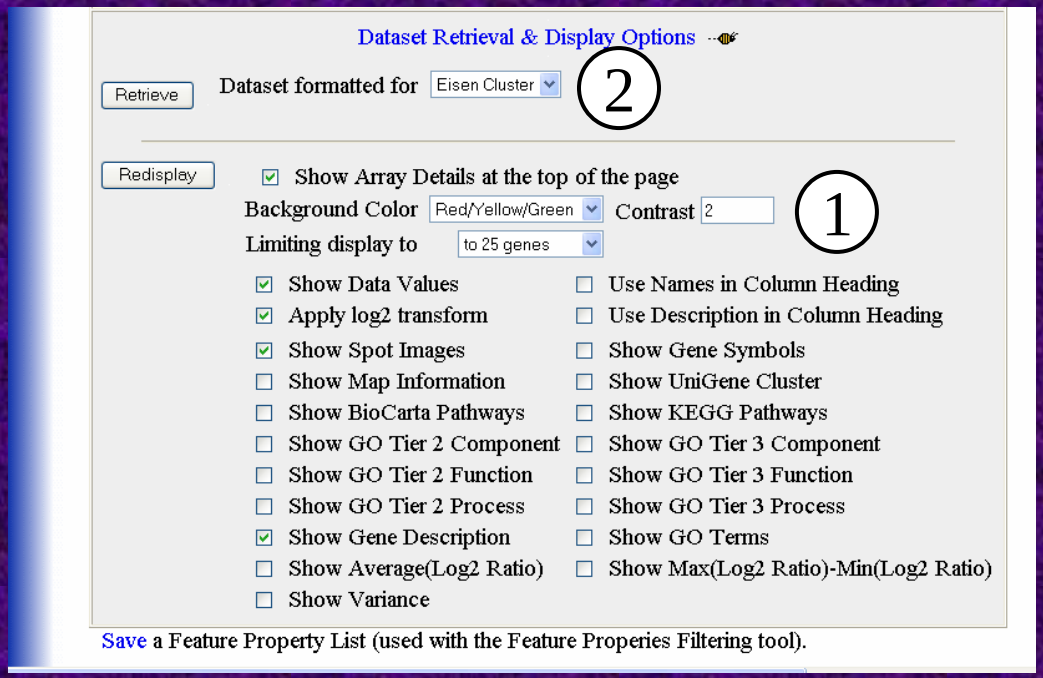
Lab 1. Main mAdb Dataset Display – Part 2



1. This is the main page to display expression data, and as we will see on the next page, is highly customizable. Each column represents an array, each row a gene feature. Gray boxes are either missing values or data that was filtered out due to low quality. You can page through the data using the **arrow** just above the columns of data.
2. The mAdb **Well ID** uniquely identifies the piece of DNA used on that feature, and the **Feature ID** is an external identifier. The **Well ID** is a hyperlink to a montage of the spot images and raw signal values, whereas the **Feature ID** is a Hyperlink to a **Feature Report**, integrating information about the gene related to the feature and its function(s).
3. There is a brief description of the feature on the right hand side of the display. Note that each column can be sorted in either ascending or descending order using the **grey arrows** above each column.


Feature Report - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Links
mAdb Feature Report
Clone IMAGE:301551
Library Source Soares_fetal_lung_NbHL19W
Sequence Verification Unknown
Annotated Simple PID Integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)
Annotated NG Assignment M14648 Human cell adhesion protein (vitronectin) receptor alpha subunit mRNA, complete cds
Annotated Categories Adhesion
5' Sequence W17002 UCSC's GenomeViewer
5' UG Title integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)
5' UG Cluster TP Hs.295726 NCBI's LocusLink Stanford's S.O.U.R.C.E.
5' UG Gene ITGAV GeneCards MedMiner NCBI's Map Viewer
5' UG LL Summary ITAGV encodes integrin alpha chain V. Integrins are heterodimeric integral membrane proteins composed of an alpha chain and a beta chain. The I-domain containing integrin alpha V undergoes post-translational cleavage to yield disulfide-linked heavy and light chains, that combine with multiple integrin beta chains to form different integrins. Among the known associating beta chains (beta chains 1,3,5,6, and 8; 'ITGB1', 'ITGB3', 'ITGB5', 'ITGB6', and 'ITGB8'), each can interact with extracellular matrix ligands; the alpha V beta 3 integrin, perhaps the most studied of these, is referred to as the Vitronectin receptor (VNR). In addition to adhesion, many integrins are known to facilitate signal transduction.
5' UG Ontology
GO ™ Annotations Evidence Source Pub
cell adhesion P Proteome PM
cell adhesion receptor P Proteome PM
integral plasma membrane protein P Proteome PM
Other Annotations
Integral membrane NR Proteome PM
Receptor (signalling) NR Proteome PM
Control of Cell Proliferation E Proteome PM
5' UG RefSeq NM_002210
5' UG Cytoband 2q31-q32
5' Submitted PID gb:M14648 VITRONECTIN RECEPTOR ALPHA SUBUNIT PRECURSOR (HUMAN);
Internet

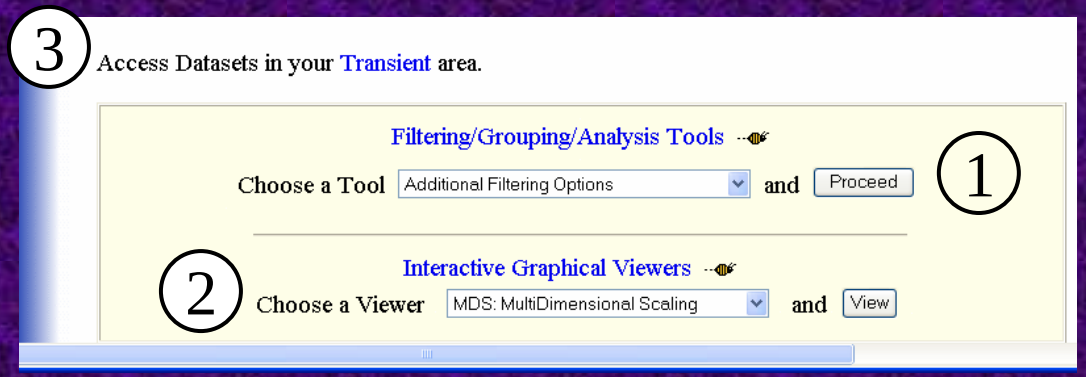
Lab 1. Main mAdb Dataset Display – Part 3



1. Here is where the data display on the preceding page can be customized, by checking or unchecking the checkboxes next to each column name. One can include numerical summary data (**Average(Log2 Ratio), Variance, Max(Log2 Ratio)-Min(log2 Ratio);** pathways (**KEGG, BioCarta**); Genome Ontology (**GO**) classifications; and display individual **Spot Images**, among others. One can also change or eliminate the **Background Color** on the table of data values, adjust its **Contrast** (the point where max red and green are reached), and also adjust how many genes are displayed in the table on a Web page (the default is 25). Once the choices are made, push the **Redisplay** button to refresh the page with your desired changes.

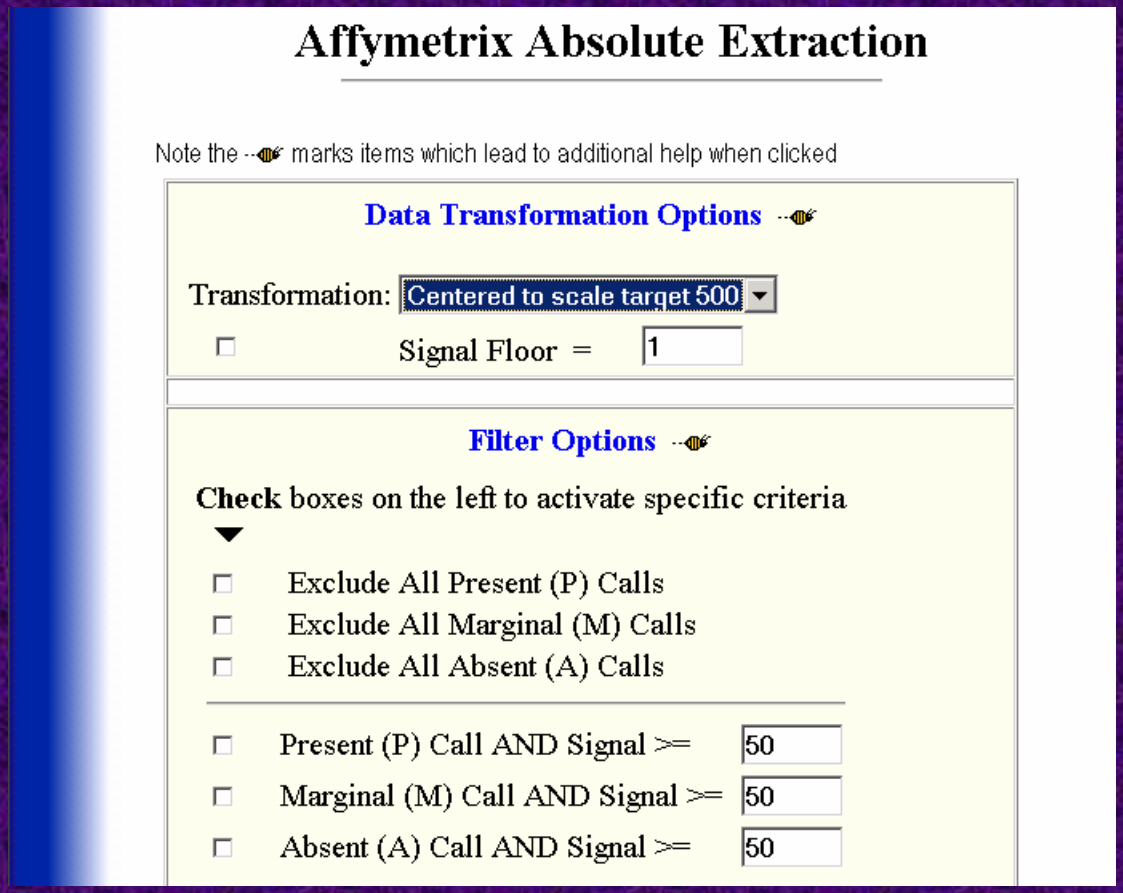
2. You can also retrieve the dataset for MS-Excel, the Eisen Cluster program format, or in tab-delimited files for the Macintosh, PC, or UNIX platforms.

Lab 1. Main mAdb Dataset Display – Part 4



1. Once the data is filtered by quality, the most likely next step is to do additional filtering and create a subset of this parent dataset. Under *Filtering/Grouping/Analysis Tools*, choose the default pulldown option of **Additional Filtering Options** and press **Proceed.**
2. Alternately, one could access *Interactive Graphical Viewers* from here,
3. Also, you could **Access** other **Datasets in your Transient Area** from here with the link above the yellow panels.

Affy Extraction Tool (for Absolute data)



Break

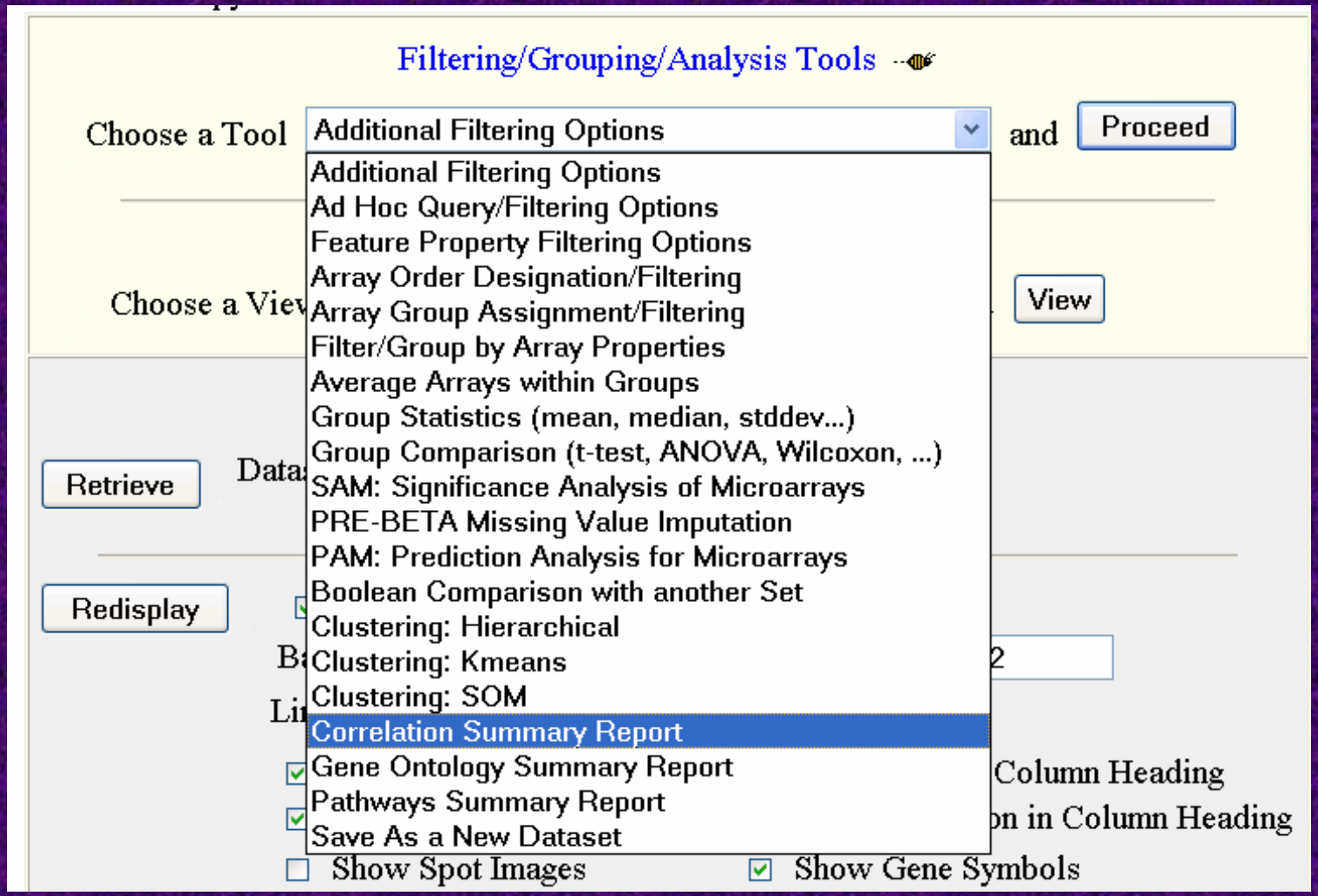
Sample Analysis Questions

- How can I evaluate the consistency of the arrays across my biological repeats?
- Which genes have enough data points to give confidence in the results?
- Which genes have values that are less consistent across the arrays?
- How can I keep track of these genes that seem to have unreliable values?
- Which genes are most differentially expressed?
- Are any of these genes in my “unreliable” list?

Lab 2 – Assessing array correlation

Goal: To evaluate the consistency of data values across a set of arrays and determine which genes are not well correlated based on a minimal number of data points

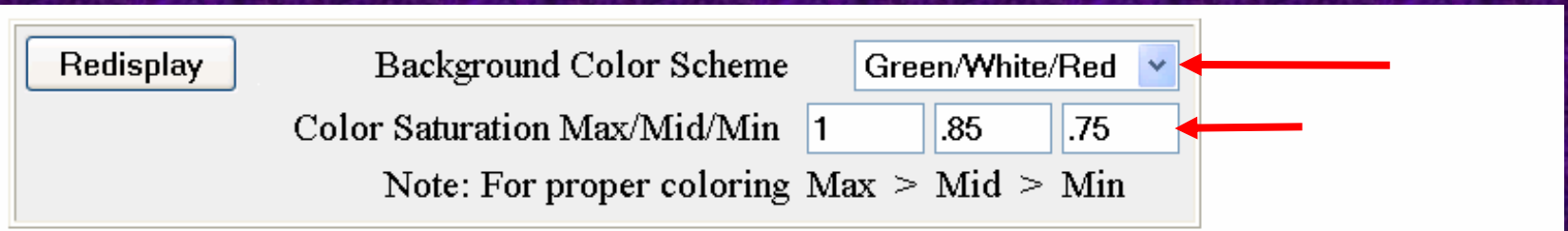
Evaluating correlation across all pairs of arrays



From the mAdb Dataset Display Page, select the "Correlation Summary Report" Tool and hit the "Proceed" button

Correlation Summary Report

(How can I evaluate the consistency of the arrays across my biological repeats?)



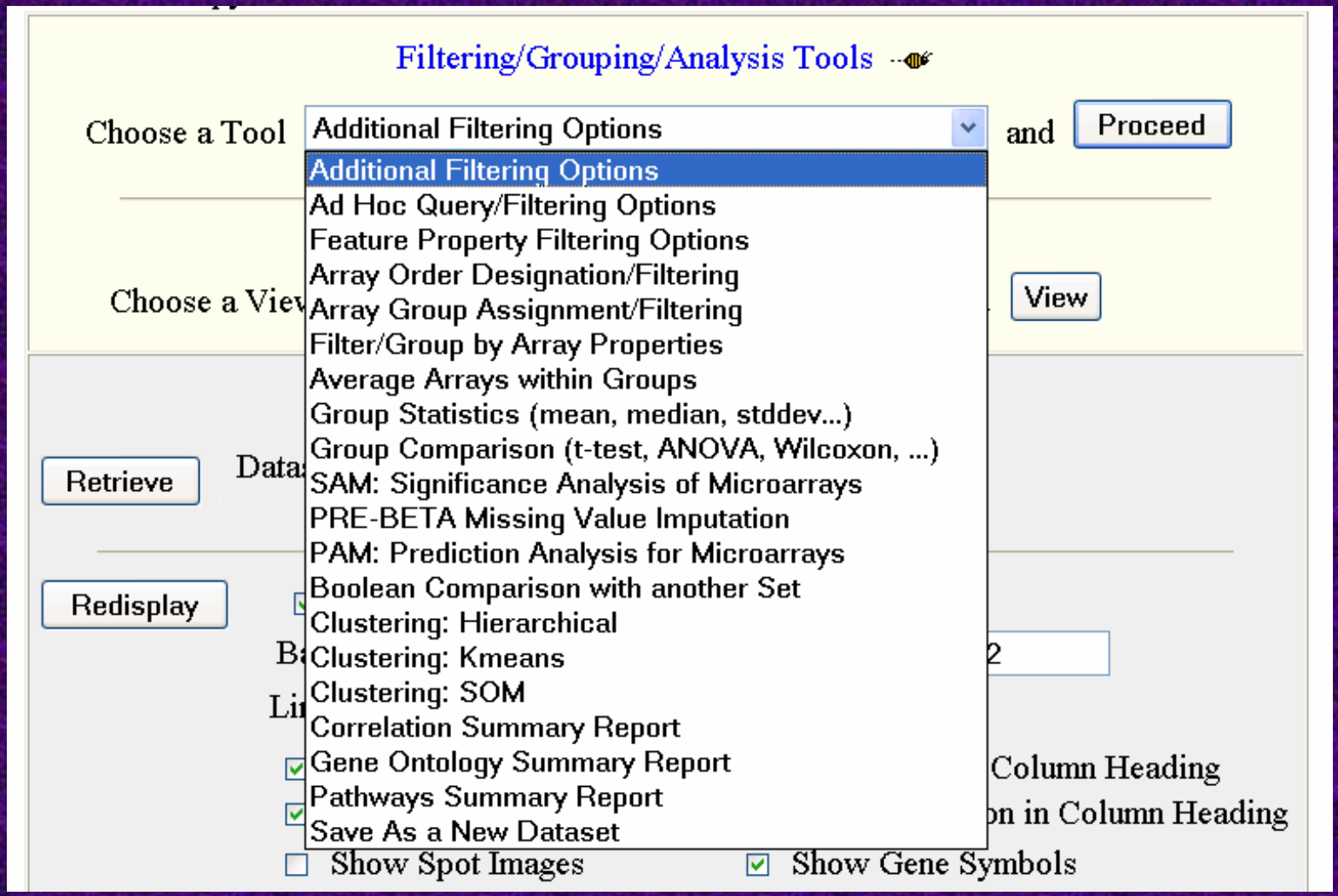
Note: Click on the Correlation values to display the corresponding ScatterPlot

Correlations

	A	A	A	A	A				
	#1	#2	#3	#4	#5	Grp		Array Name	Array Description
1.	#1 A	0.855	0.927	0.917	0.912	A	1.	Mm-Incyte-v1p1-1	Sample 1/Type A
2.		#2 A	0.802	0.844	0.831	A	2.	Mm-Incyte-v1p1-2	Sample 2/Type A
3.			#3 A	0.948	0.935	A	3.	Mm-Incyte-v1p1-3	Sample 3/Type A
4.				#4 A	0.940	A	4.	Mm-Incyte-v1p1-4	Sample 4/Type A
5.					#5 A	A	5.	Mm-Incyte-v1p1-5	Sample 5/Type A

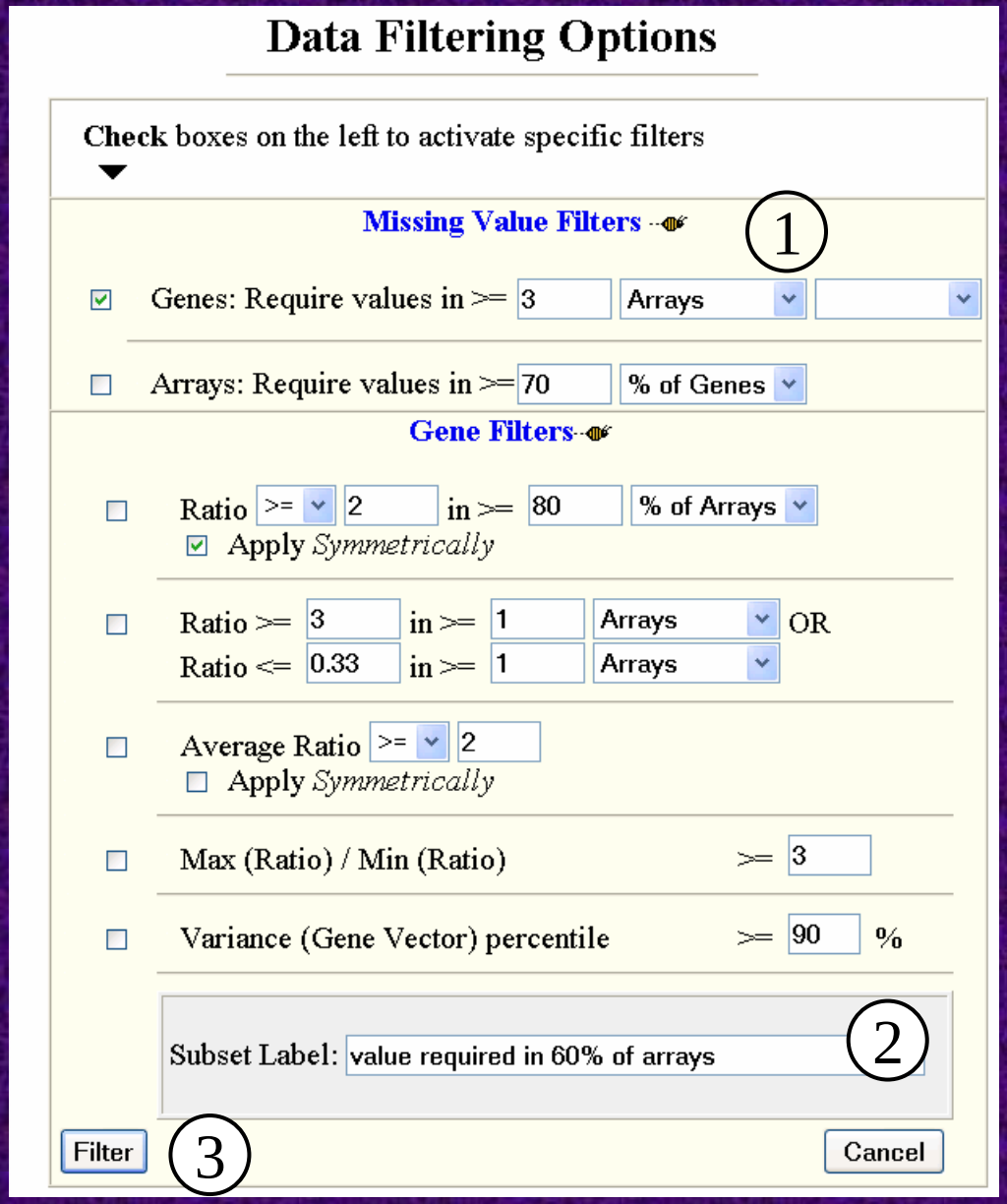
Allows pair wise comparison of all arrays in a project – useful for comparing replicates and reverse fluors

Selecting spots based on value characteristics



From the mAdb Dataset Display Page, select the "Additional Filtering Options" Tool and hit the "Proceed" button

Filtering based on missing values



1. Filter the rows of data from the parent dataset for missing values, requiring genes in **>=3** Arrays. Alternately, it is possible to filter out Arrays by requiring values in >= 60% of genes.
2. Label the subset "value required in 60% of arrays"
3. Press the **Filter** button to continue and create the desired subset.

Filtering based on missing values

(Which genes have enough data points to give confidence in the results?)

Edit Data for Subset: **value required in 60% of arrays**
from Dataset: **Extracted type A**

1

```
The filter input data set contained 5 arrays and 7525 genes.
The filtered output data set contains 5 arrays and 3771 genes.
3754 genes excluded for being present in less than 60% (3) arrays.
```

View the complete History. 2

Expand this Dataset.
Access Datasets in your Temporary area.

3

Records 1 to 25 of 3771 total records displayed.

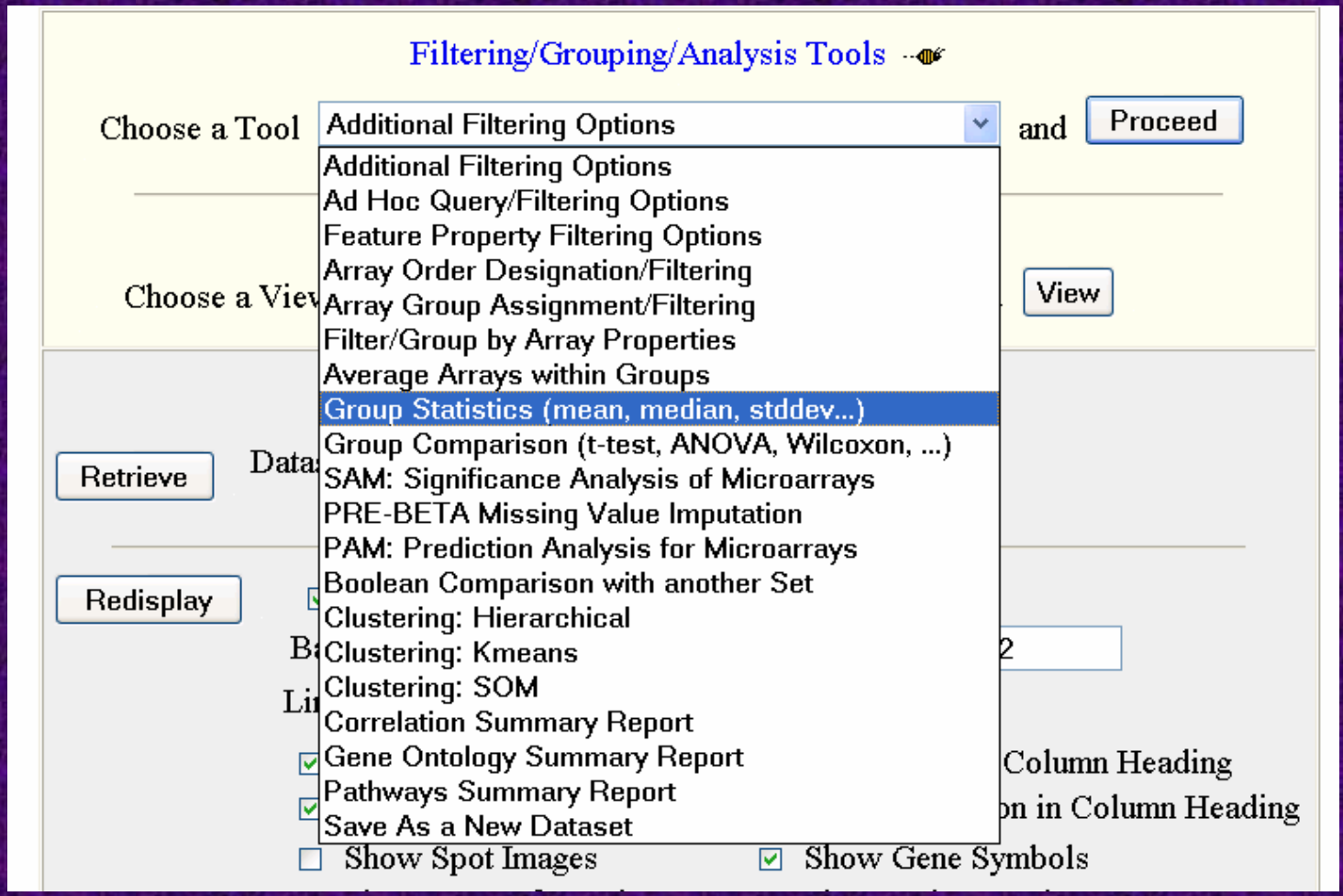
A	A	A	A	A			
#1	#2	#3	#4	#5	Well ID	Feature ID	Ge
-0.5580	-0.5632	-0.1758	-0.4063	-0.4641	621790	IMAGE:651430	Mkrn2
	-0.5094	-0.6059		-0.2902	621785	IMAGE:523795	Ptdss1
0.5718	-0.0435	0.4337	0.6030	0.4537	621781	IMAGE:533862	Ccl3
	0.3440	0.5024		0.2958	621779	IMAGE:533299	Xpo1
	2.7681	1.9809		1.4138	621777	IMAGE:522713	

1. Note that in the returned dataset, there are many fewer missing values – see the history log for how many genes were filtered out to create this subset.
2. This is a data subset – you can view the complete History of the dataset via this link.
3. You can also **Expand this Dataset** to show the parent and all children, or again **Access Datasets in your Temporary Area** via these links.

Notes:

- Applies selected filtering options to the dataset based on values in the data and creates a new subset.
- For gene filters, ratios are expressed as fold changes and all calculations are done in log space

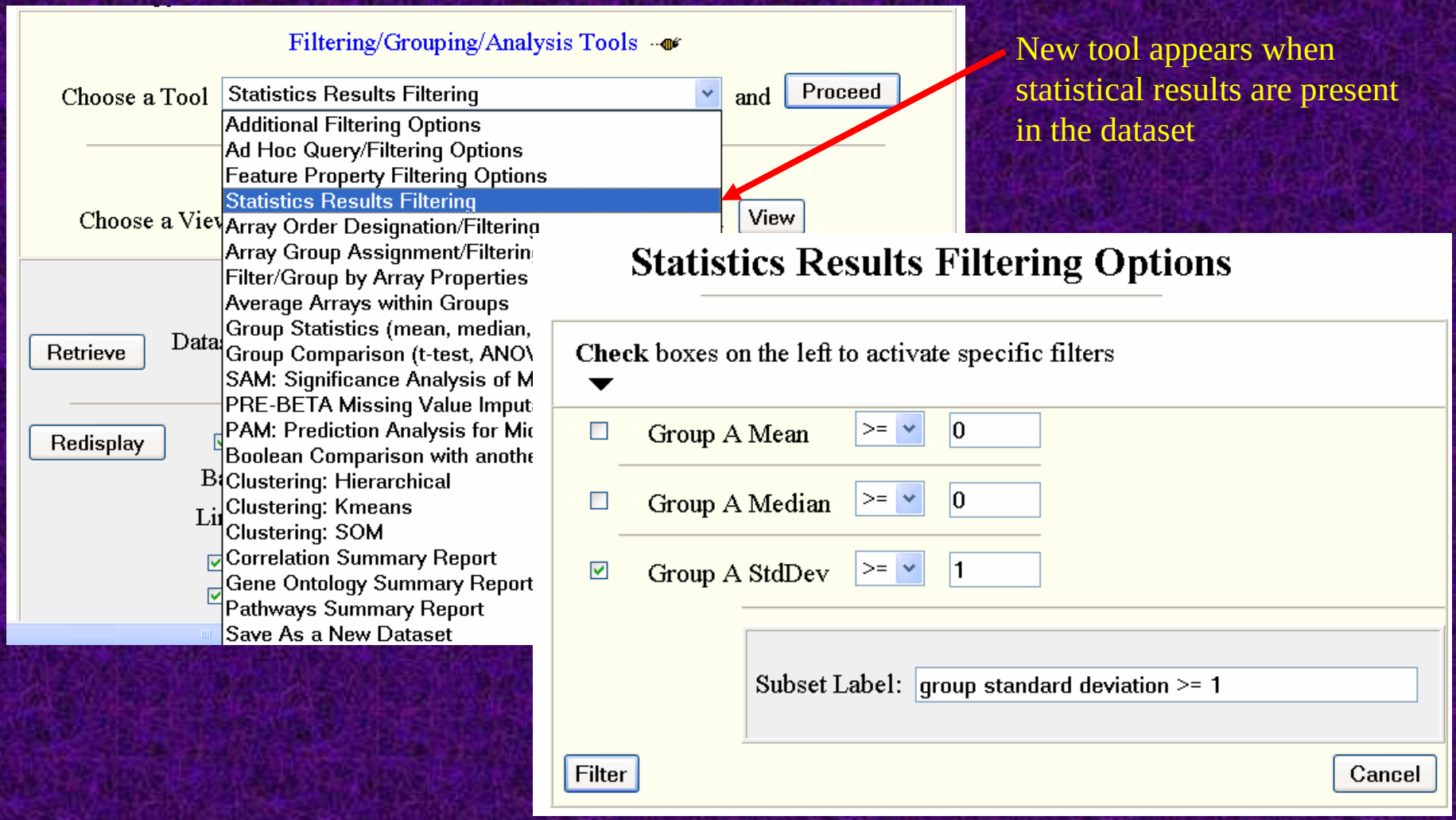
Calculating Group Statistics



From the mAdb Dataset Display Page, select the "Group Statistics" Tool and hit the "Proceed" button

Filtering on Group Statistics

(Which genes have values that are less consistent across the arrays?)



Sorting on Group Statistics

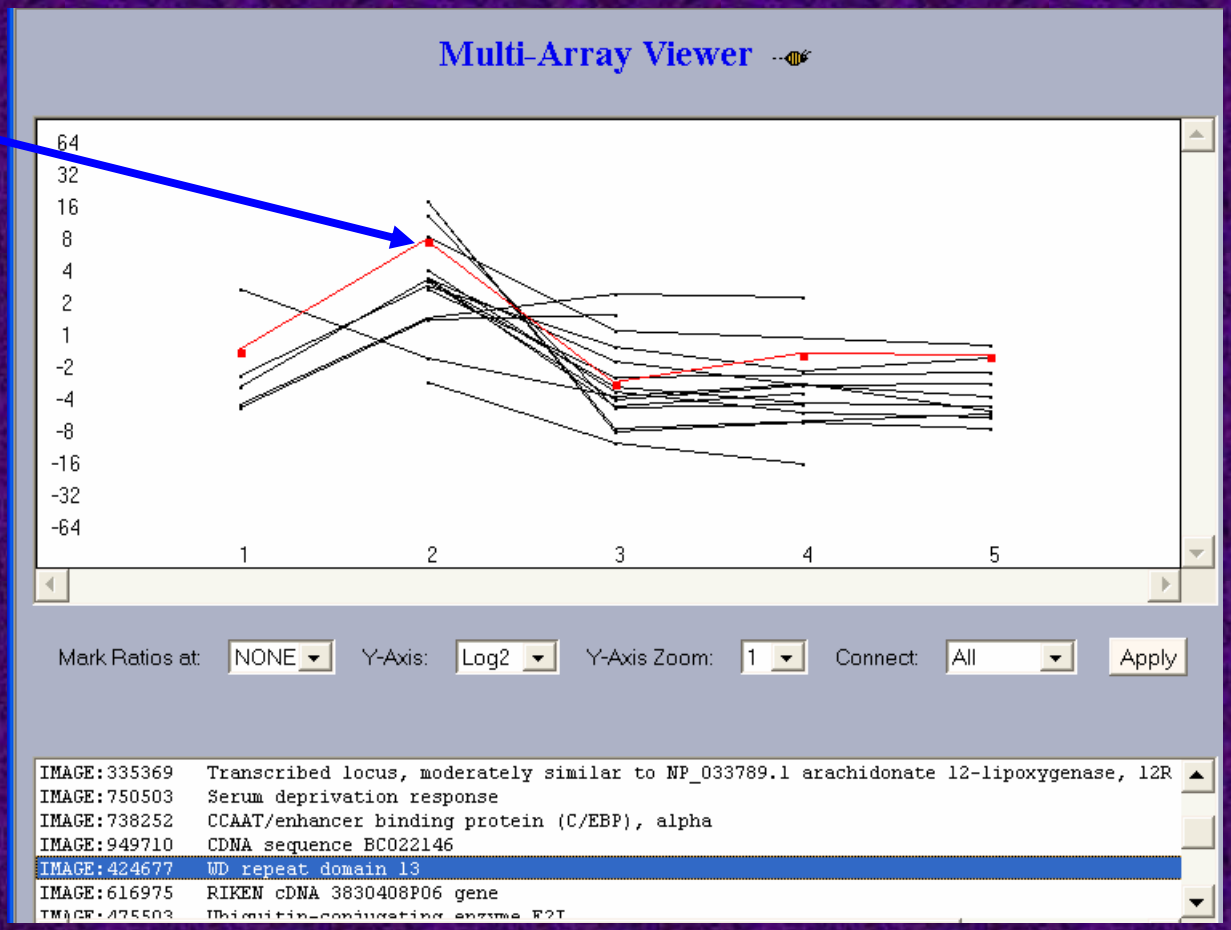
☑ Group Means ☐ Group Medians
☑ Group StdDevs

Save a Feature Property List (used with the Feature Properties Filtering tool).

Records 1 to 16 of 16 total records displayed.

User can sort rows by clicking on up/down arrows above columns

A	A	A	A	A					
#1	#2	#3	#4	#5	Group A Mean	Group A StdDev	Well ID	Feature ID	Gene
	4.4465	-2.6394		-2.2019	-0.1316	3.242	616653	IMAGE:480196	
	4.4261	-2.7721	-2.4251	-2.6466	-0.8544	3.051	613650	IMAGE:336497	Slc39a4
	3.9956	-2.0067	-1.9103		0.0262	2.807	613408	IMAGE:331681	Pcbd
	2.3167	-1.5024	-2.1655	-2.3175	-0.9172	1.892	621404	IMAGE:777580	
	2.0528	-1.9512	-1.5359		-0.4781	1.798	620614	IMAGE:738252	Cebpa
-0.1835	3.2536	-1.2083	-0.2851	-0.3586	0.2436	1.549	615066	IMAGE:424677	Wdr13
	3.3436	0.3888		-0.0690	1.2212	1.512	618887	IMAGE:643725	Ramp2
	1.9644	-1.7476	-1.2831	-1.2732	-0.5849	1.484	613517	IMAGE:330336	H2-K1
-1.3422	2.0174	-1.3657	-1.8360	-1.9434	-0.8940	1.476	621175	IMAGE:803488	Sfrp1
	2.0624	-0.5530	-1.2824	-1.6303	-0.3508	1.447	621320	IMAGE:790857	Tex261
-1.9145	0.8009	1.5663	1.4701		0.4807	1.414	616332	IMAGE:475503	Ube2i
1.6904	-0.4691	-1.6724	-1.2459	-2.0954	-0.7585	1.337	619489	IMAGE:949710	BC022146
-2.0005	0.7449	0.9103			-0.1151	1.335	617168	IMAGE:573075	Bag2
	1.6819	-1.0463		-0.8763	-0.0802	1.248	613545	IMAGE:335369	
	-1.2203	-3.0920	-3.7503		-2.6875	1.072	620919	IMAGE:750503	Sdpr
-1.0227	1.8079	-0.1077	-0.8707	-0.4519	-0.1290	1.02	617980	IMAGE:616975	3830408P0

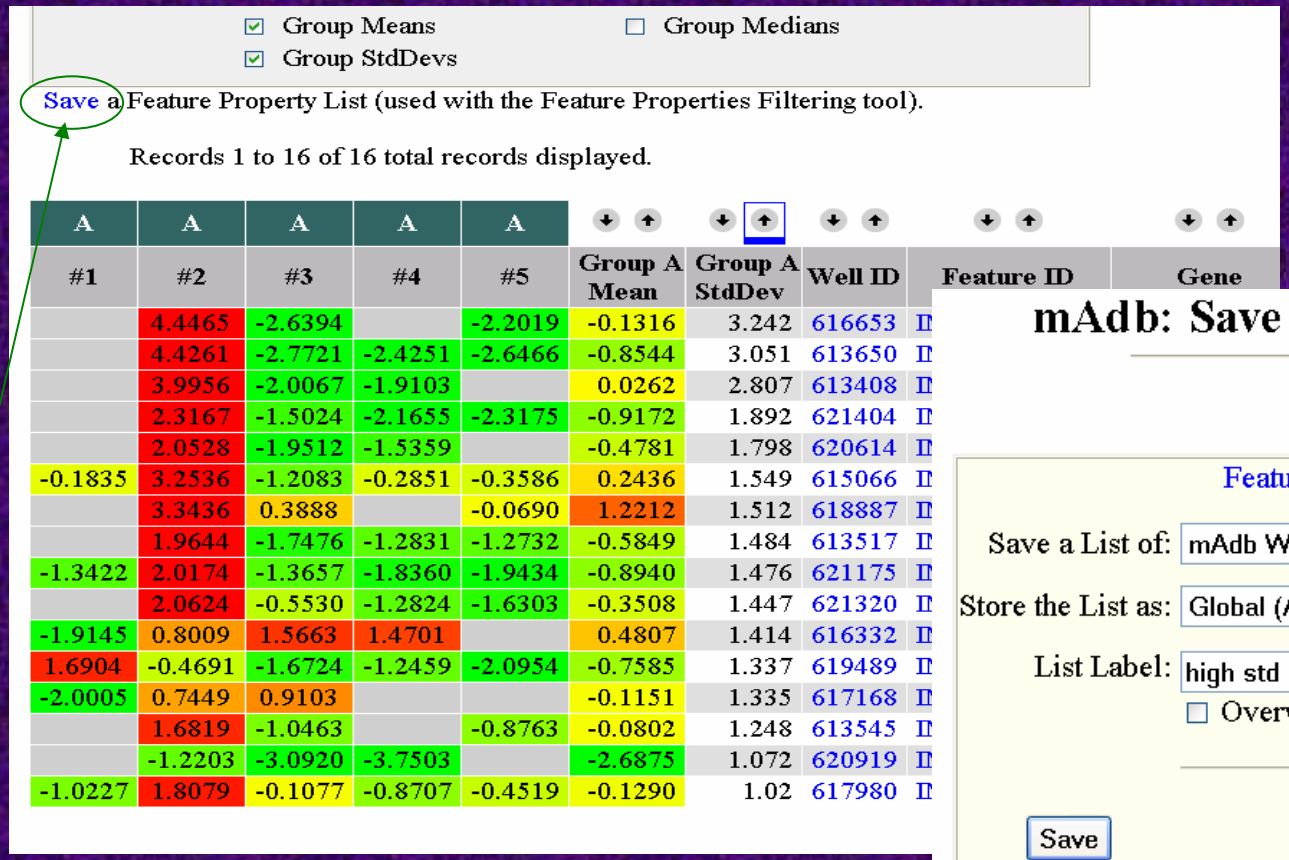


Access from *Interactive Graphical Viewers* Menu on main **mAdb Dataset Display** page :

1. Can choose a point on graphical window to display a graph of that gene's expression which passes through that point
2. Can select a gene name on lower list and graph will appear in plot above
3. Can get **Feature Report** by clicking on gene name in lower display box

Save a Feature Property List

(How can I keep track of these genes that seem to have unreliable values?)



☑ Group Means ☐ Group Medians
☑ Group StdDevs

Save a Feature Property List (used with the Feature Properties Filtering tool).

Records 1 to 16 of 16 total records displayed.

A	A	A	A	A					
#1	#2	#3	#4	#5	Group A Mean	Group A StdDev	Well ID	Feature ID	Gene
	4.4465	-2.6394		-2.2019	-0.1316	3.242	616653		
	4.4261	-2.7721	-2.4251	-2.6466	-0.8544	3.051	613650		
	3.9956	-2.0067	-1.9103		0.0262	2.807	613408		
	2.3167	-1.5024	-2.1655	-2.3175	-0.9172	1.892	621404		
	2.0528	-1.9512	-1.5359		-0.4781	1.798	620614		
-0.1835	3.2536	-1.2083	-0.2851	-0.3586	0.2436	1.549	615066		
	3.3436	0.3888		-0.0690	1.2212	1.512	618887		
	1.9644	-1.7476	-1.2831	-1.2732	-0.5849	1.484	613517		
-1.3422	2.0174	-1.3657	-1.8360	-1.9434	-0.8940	1.476	621175		
	2.0624	-0.5530	-1.2824	-1.6303	-0.3508	1.447	621320		
-1.9145	0.8009	1.5663	1.4701		0.4807	1.414	616332		
1.6904	-0.4691	-1.6724	-1.2459	-2.0954	-0.7585	1.337	619489		
-2.0005	0.7449	0.9103			-0.1151	1.335	617168		
	1.6819	-1.0463		-0.8763	-0.0802	1.248	613545		
	-1.2203	-3.0920	-3.7503		-2.6875	1.072	620919		
-1.0227	1.8079	-0.1077	-0.8707	-0.4519	-0.1290	1.02	617980		

mAdb: Save a Feature Property List

Feature Property List

Save a List of: mAdb Well IDs

Store the List as: Global (Available in all Datasets)

List Label: high std dev genes

☐ Overwrite any existing list with the same label

Save

- Can save a list of well IDs, clone/feature identifiers, gene symbols, UniGene identifiers from the dataset display page
- List can be stored as local to the dataset or globally available to all datasets

Dataset History

History for Subset: **group standard deviation >= 1**
from Dataset: **Extracted type A**

```
5 Arrays and 7525 Expression Rows extracted.
Default Ratio: ChanB/ChanA (Cy5/Cy3)
Signal calculation: Median Intensity minus Median Background
Any Features designated Control were excluded.
Normalization method: 50th Percentile (Median) using all spot filtered Genes
Spot Filter Options:
Include Spots not flagged BAD or Not Found
AND Both Chan A and Chan B Signal >= 200
Override other Chan A & B criteria and Include if Chan A Signal >= 5000 OR Chan B Signal >= 5000
Data was extracted and aligned by the Inventory Well ID
Any multiple occurrences of Well ID were reduced to a single instance
by selecting the one with the strongest signal (Chan A + Chan B)
Note: For all GenePix results from Axon scanned arrays Chan A is CY3 and Chan B is CY5.
Rows ordered by Average(Log Ratio) descending.
------------------------------------------------------------
Fri Aug 19 11:29:03 EDT 2005

5 arrays, 7525 genes in the original Dataset
3771 Genes and 5 arrays passed filters
3754 genes excluded for being present in less than 60% (3) arrays.

------------------------------------------------------------
Fri Aug 19 11:29:35 EDT 2005

input Dataset
Group Statistic calculations performed for each Group
------------------------------------------------------------
Fri Aug 19 11:35:07 EDT 2005

3771 genes in the input Dataset
The filtered output data set contains 16 genes
3755 genes excluded by Group A StdDev >= 1
```

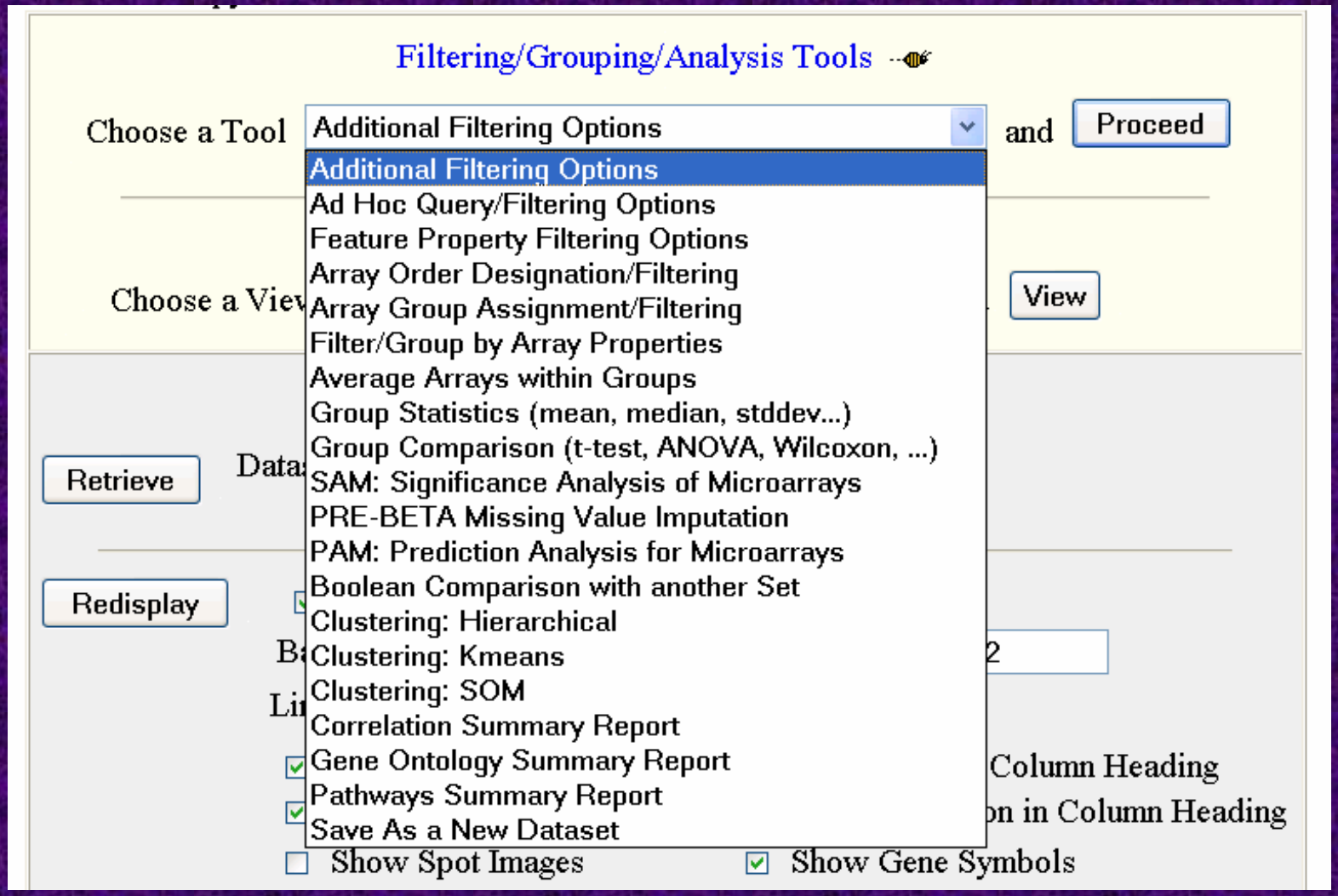
Link to the output Dataset

A log is maintained for each dataset tracing the analysis history. When the history is displayed, links are provided to allow the user to recall any dataset in the analysis chain.

Lab 3 – Examining differentially expressed genes

Goal: To find differentially expressed genes and evaluate the reliability of values

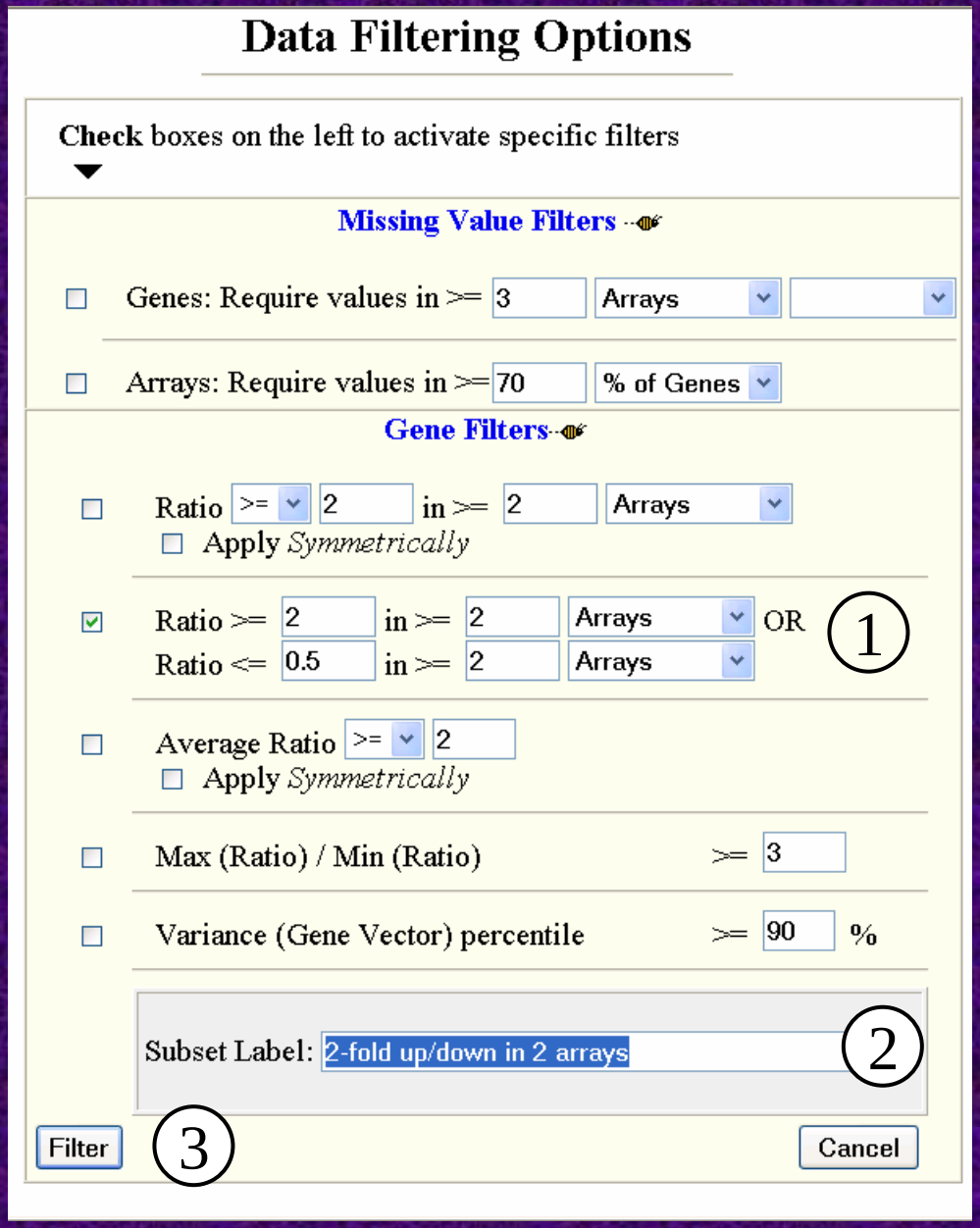
Refining spot selection criteria



From the mAdb Dataset Display Page, select the "Additional Filtering Options" Tool and hit the "Proceed" button

Filtering on data values

(Which genes are most differentially expressed?)



1. Filter for at least 2-fold up in 2 or more arrays OR at least 2-fold down in 2 or more arrays.

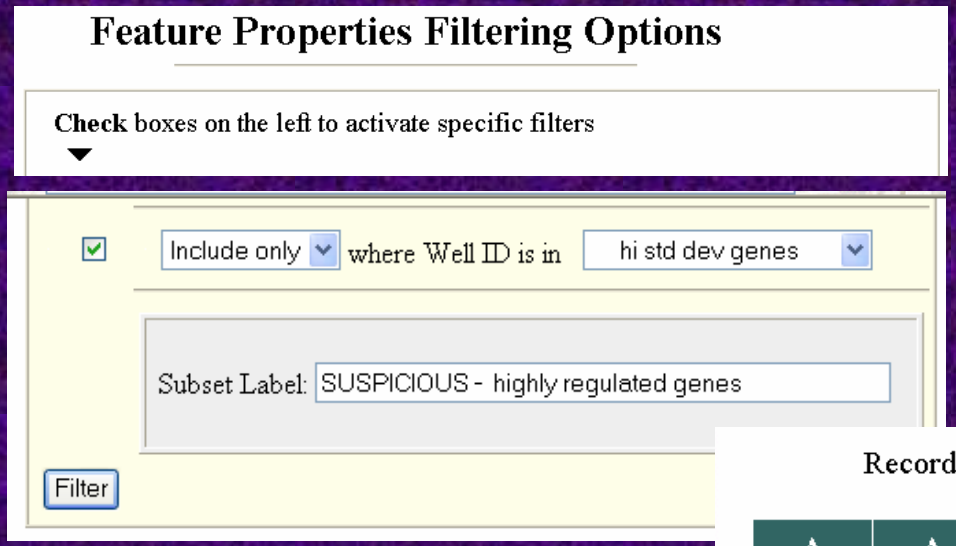
   Other options are:

   - Filter **Ratio >= 2 in >=2 Arrays**, with the **Apply *Symmetrically*** box checked to obtain genes up or down-regulated by 2-fold or more.
   - Filter for an average Ratio across the row at least two fold or more, applied symmetrically to obtain genes with an average ratio two-fold or more up or down regulated.
   - Filter for those rows showing a difference between the maximum ratio and minimum ratio on each row of 2 fold or more
   - Rank the genes by percentile of variance, and then filter for those genes in the top 10%ile of variance – ie. The genes that vary the most across the rows statistically.
   - N.B. Filters are applied in order from top to bottom – can iteratively access this tool to filter in your preferred order

2. Label the subset "2-fold up/down in 2 arrays"
3. Press the **Filter** button to continue and create the desired subset.

Filtering by Feature Properties and/or Lists

(Are any of these genes in my “unreliable” list?)



Filters any dataset so that only those identifiers matching feature properties in the selected list are included (or excluded)

Records 1 to 11 of 11 total records displayed.

A	A	A	A	A			
#1	#2	#3	#4	#5	Well ID	Feature ID	Gene
-1.9145	0.8009	1.5663	1.4701		616332	IMAGE:475503	Ube2i
	3.9956	-2.0067	-1.9103		613408	IMAGE:331681	Pcbd
	4.4465	-2.6394		-2.2019	616653	IMAGE:480196	
	2.0624	-0.5530	-1.2824	-1.6303	621320	IMAGE:790857	Tex261
	2.0528	-1.9512	-1.5359		620614	IMAGE:738252	Cebpa
	1.9644	-1.7476	-1.2831	-1.2732	613517	IMAGE:330336	H2-K1
1.6904	-0.4691	-1.6724	-1.2459	-2.0954	619489	IMAGE:949710	BC022146
	4.4261	-2.7721	-2.4251	-2.6466	613650	IMAGE:336497	Slc39a4
-1.3422	2.0174	-1.3657	-1.8360	-1.9434	621175	IMAGE:803488	Sfrp1
	2.3167	-1.5024	-2.1655	-2.3175	621404	IMAGE:777580	
	-1.2203	-3.0920	-3.7503		620919	IMAGE:750503	Sdpr

More analysis tools

Pathway Summary Report

Total number of features: 97
Total number of features mapped to a KEGG Pathway: 8
Total number of features mapped to a BioCarta Pathway: 5
Total number of features not mapped to any Pathway: 84

NOTE: Clicking on # of features creates a new subset containing only the features the mapped to the Pathway.
NOTE: Clicking on BioCarta Pathway ID displays the pathway.

# of Features	BioCarta Pathway	
1	m_cxcr4Pathway	CXCR4 Signaling Pathway
1	m_ifngPathway	IFN gamma Signaling Pathway
1	m_keratinocytePathway	Keratinocyte Differentiation
1	m_etsPathway	METS Affect on Macrophage Differentiation
1	m_ccr5Pathway	Pertussis toxin-insensitive CCR5 Signaling in Macrophage
1	m_nktPathway	Selective Expression of Chemokine Receptors during T-cell Polarization
1	m_malatexPathway	Shuttle for Transfer of Acetyl Groups from Mitochondria to the Cytosol
1	m_th1th2Pathway	Th1/Th2 Differentiation
1	m_eea1Pathway	The Role of FYVE-finger Proteins in Vesicle Transport

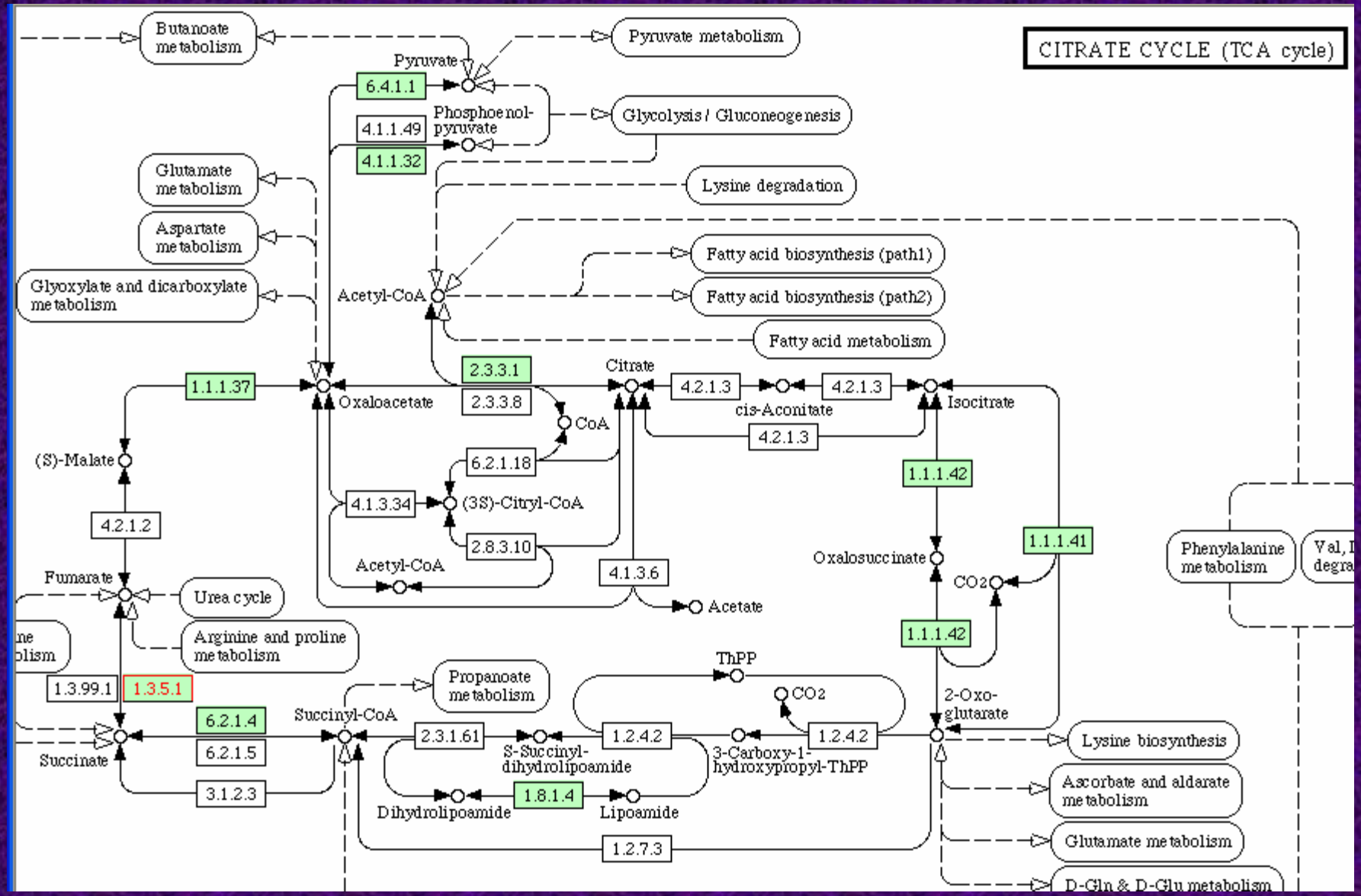
NOTE: Clicking on # of features creates a new subset containing only the features the mapped to the Pathway.
NOTE: Clicking on KEGG Pathway ID displays the pathway with features high lighted.

# of Features	KEGG Pathway	
2	mmu00561	Glycerolipid metabolism
2	mmu00190	Oxidative phosphorylation
1	mmu00193	ATP synthesis
1	mmu00362	Benzoate degradation via hydroxylation
1	mmu00710	Carbon fixation
1	mmu00020	Citrate cycle (TCA cycle)

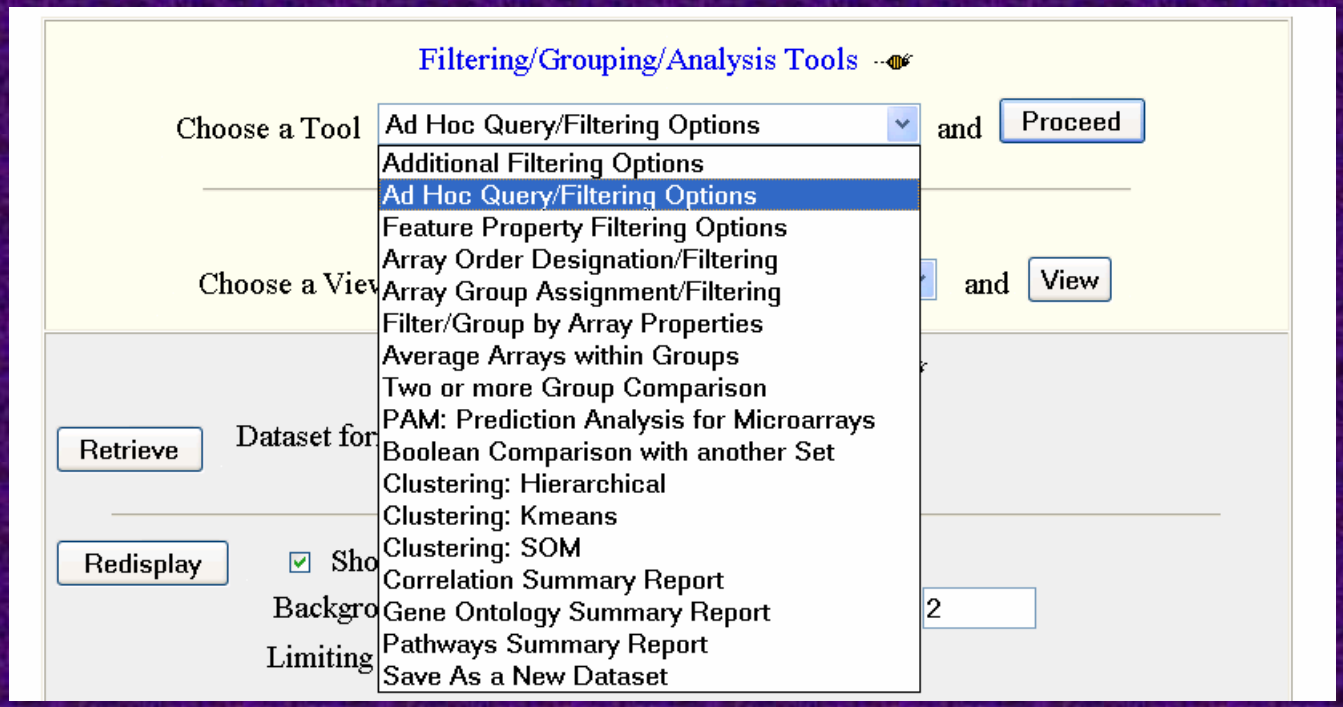
From the mAdb Dataset Display Page, select "Pathways Summary Report"

1. Clicking on **# of Features** link creates a new dataset of just those features.
2. Clicking on **BioCarta Pathway** links show pathway on BioCarta Web site.
3. GO Ontology Summary Report also available

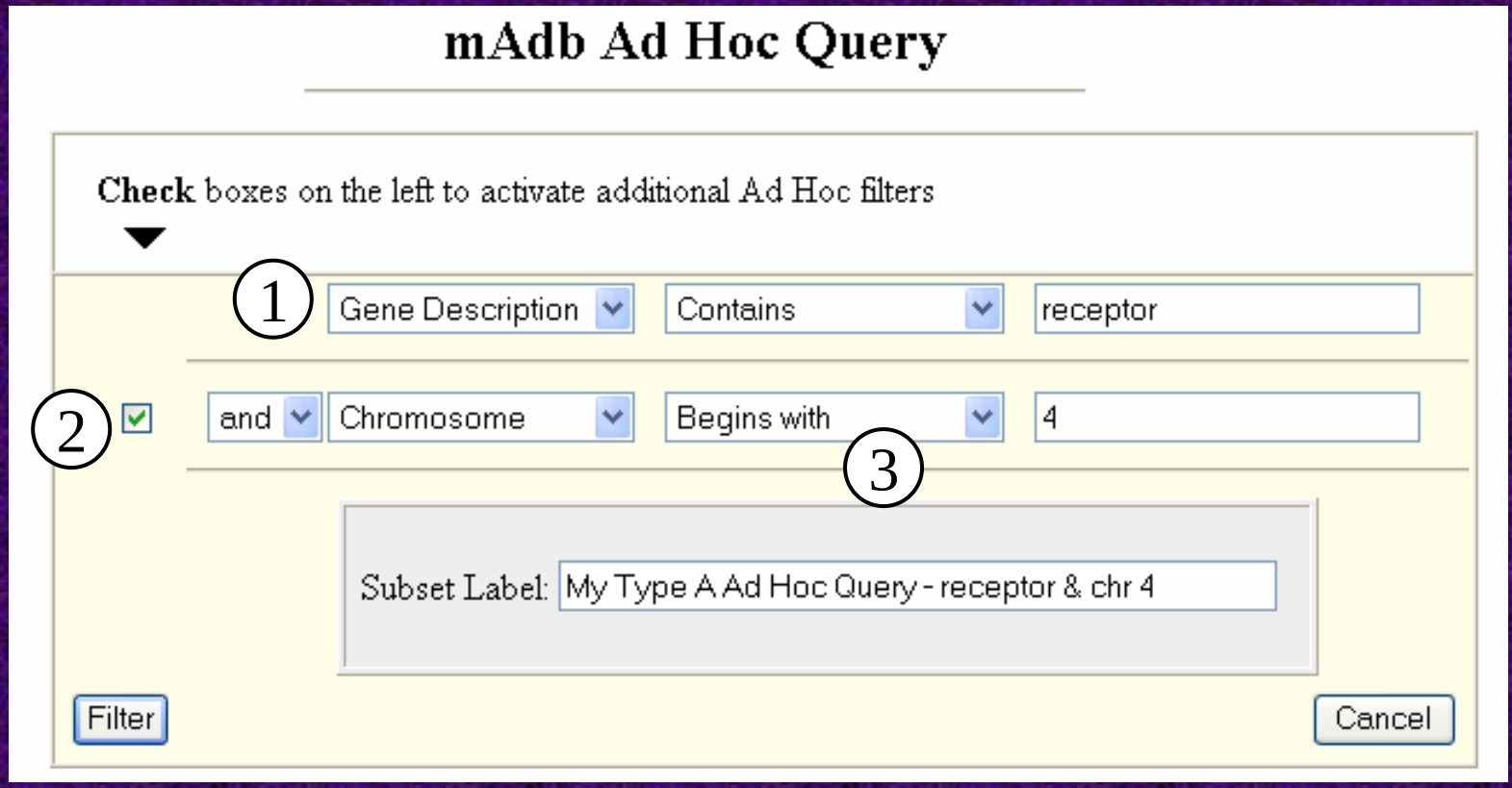
A KEGG Pathway



Ad Hoc Query Tool



From the mAdb Dataset Display Page, select the “Ad Hoc Query/ Filtering Options” Tool and hit the “Proceed” button



Boolean Keyword search.

1. Pick from **BioCarta Pathway, Feature ID, Gene Description, Gene Symbol, GO term, KEGG Pathway, Map Location, UniGene ID, Well ID category**
2. Check box to add another term with **AND/OR** choice
3. Choose **Contains, Begins With, Equals, Does Not Contain, Does Not Begin With, Does Not Equal** for search qualifier

Output of Ad Hoc Query

mAdb Dataset Display

View Array Summaries

Edit Data for Subset: **My Type A Ad Hoc Query - receptor & chr 4**
from Dataset: **test for class**

```
Ad Hoc Filtering
5 arrays and 340 genes in the input dataset
5 arrays and 2 genes in the output dataset.
Ad Hoc Filter:
    Gene Description Contains 'receptor'
AND Chromosome Begins with '4'
```

Records 1 to 2 of 2 total records displayed.

A	A	A	A	A					
#1	#2	#3	#4	#5	Aver	Well ID	Feature ID	Map	Description
3.2349	2.7053	3.0574	2.8567	3.3126	3.0334	614354	IMAGE:403453	4 C6-D1	protein tyrosine phosphatase, receptor type, F
-1.9201	-2.4286	-1.7173	-1.9279	-1.8618	-1.9711	620446	IMAGE:735186	4 D2.3	nuclear receptor binding factor 1

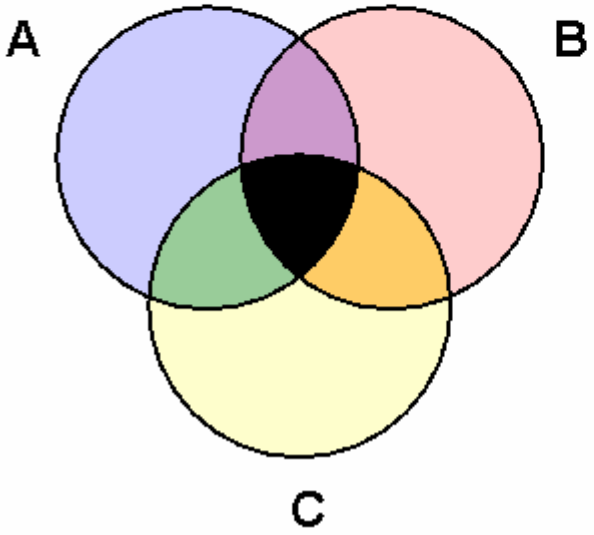
Graphical Venn Tool

Compares subset intersections

Boolean Comparison

	Label	Arrays	Genes	Created
Set A	Small, Round Blue Cell Tumors (SRBCTs),...	88	2309	Sep 18 2002 11:30:00am
Set B	PAM Threshold 3.928	63	68	Sep 20 2002 5:13:59pm
Set C	SAM Delta 0.800	10	359	Sep 10 2004 5:55:21pm

Click regions in the diagram to Select/Deselect

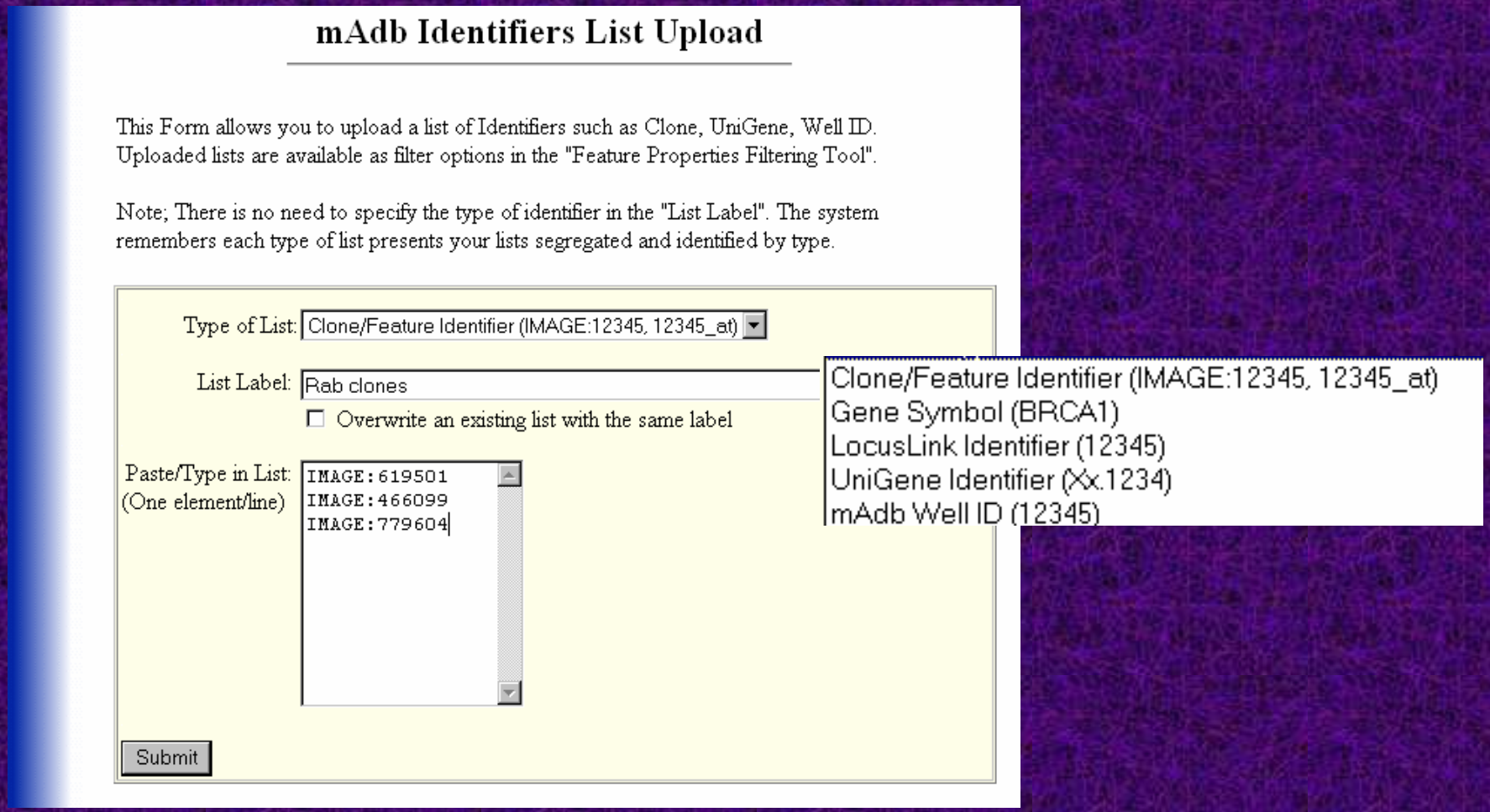


- Region 1 Abc	1923	features
- Region 2 ABc	27	features
- Region 3 Bac	0	features
- Region 4 ACb	318	features
- Region 5 ABC	**41**	**features**
- Region 6 BCa	0	features
- Region 7 Cab	0	features

Selected:	**41**	**features**

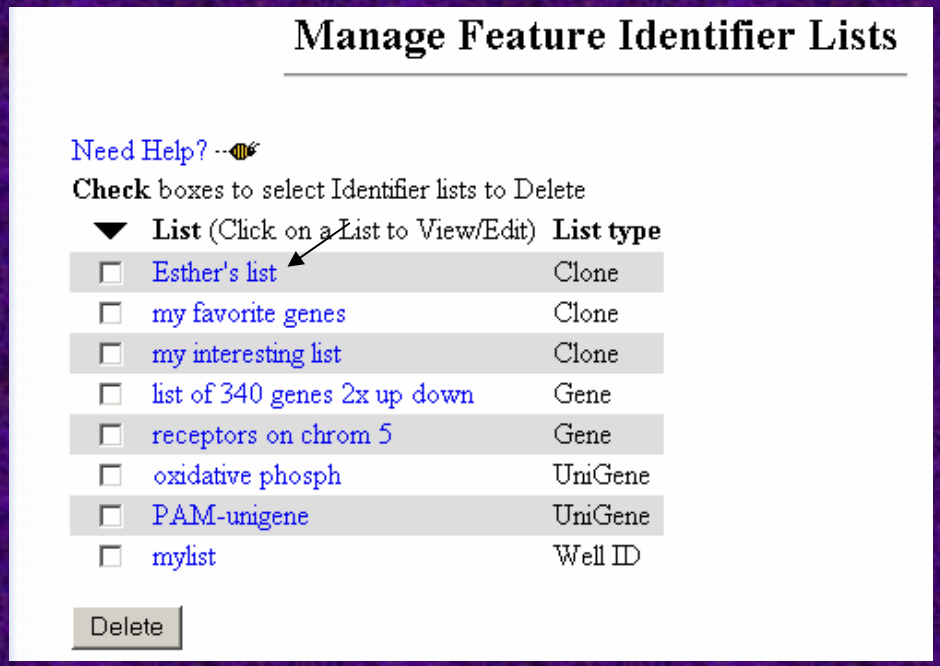
From the mAdb Dataset Display Page, select the “Boolean comparison using Venn Diagrams” Tool and hit the “Proceed” button

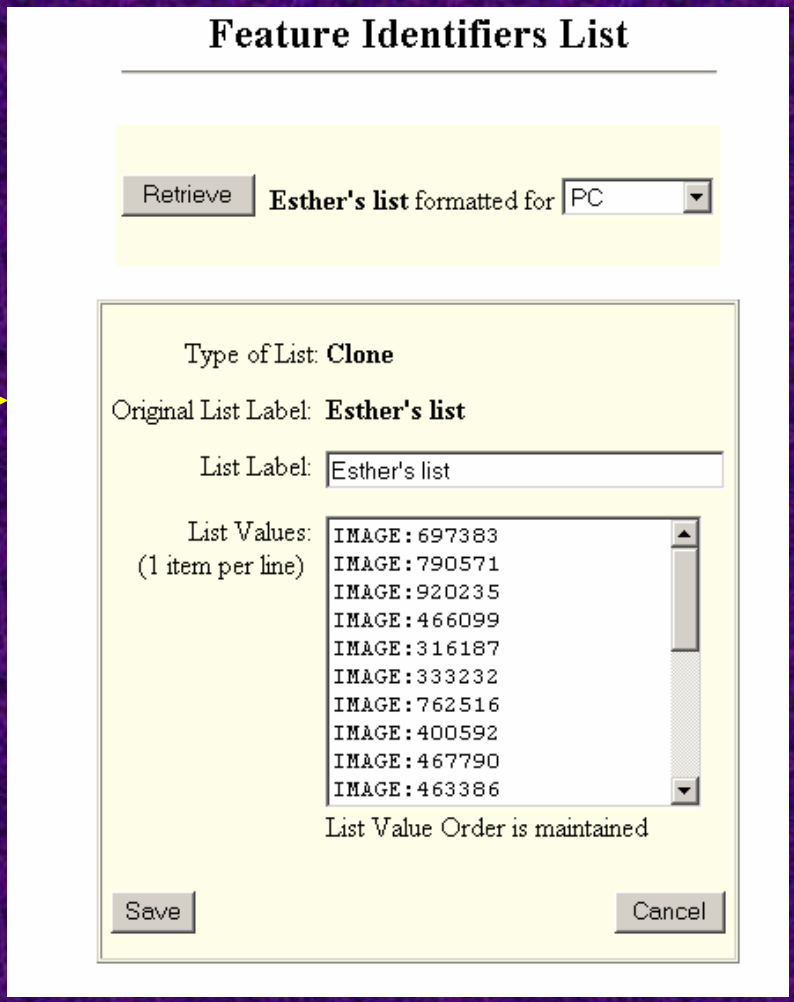
Manually Create a List of Identifiers for Filtering



From the mAdb Gateway page, use the “Upload Identifier list” link.
Paste in list of identifier (use format as shown for specific type)

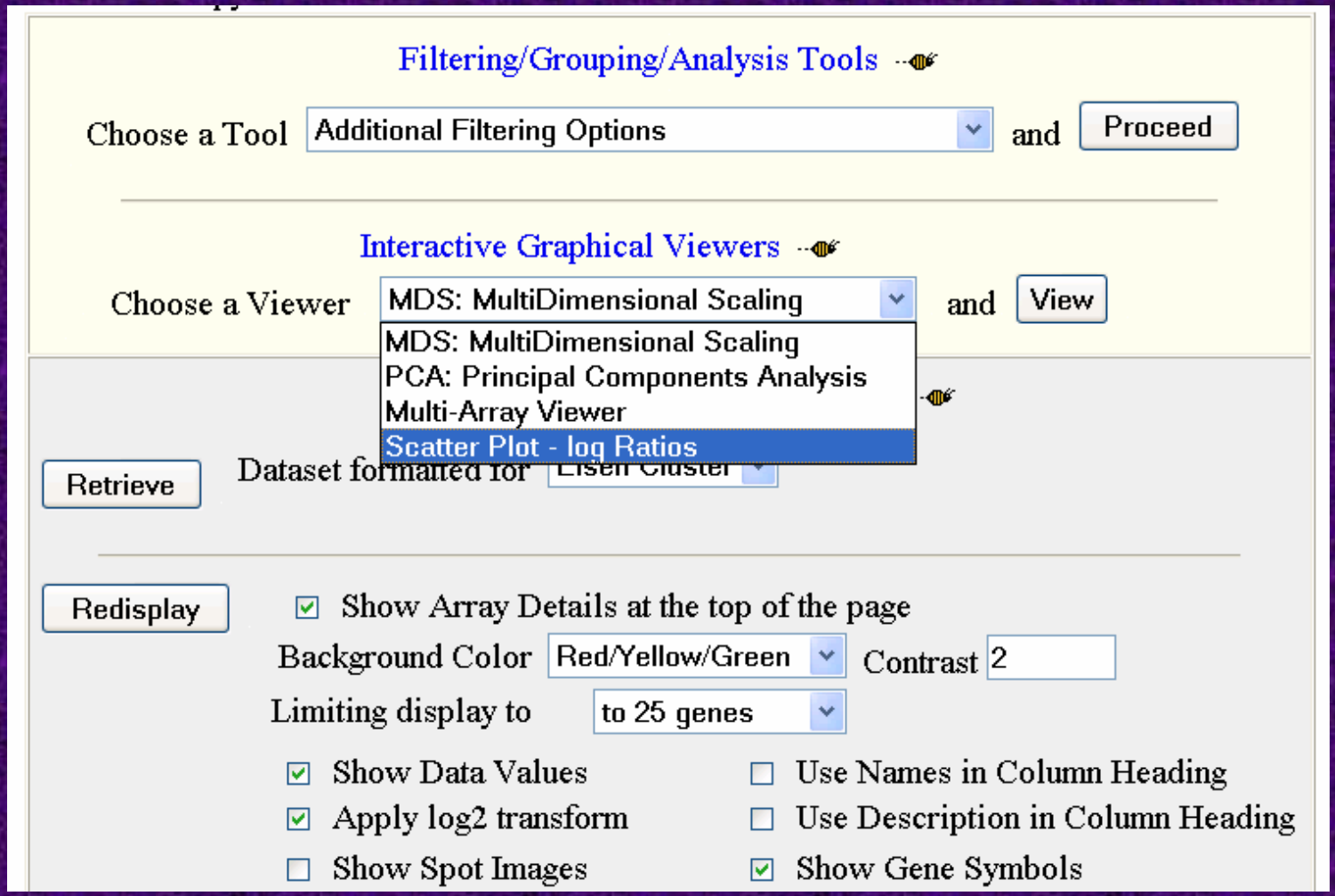
Managing Feature Lists





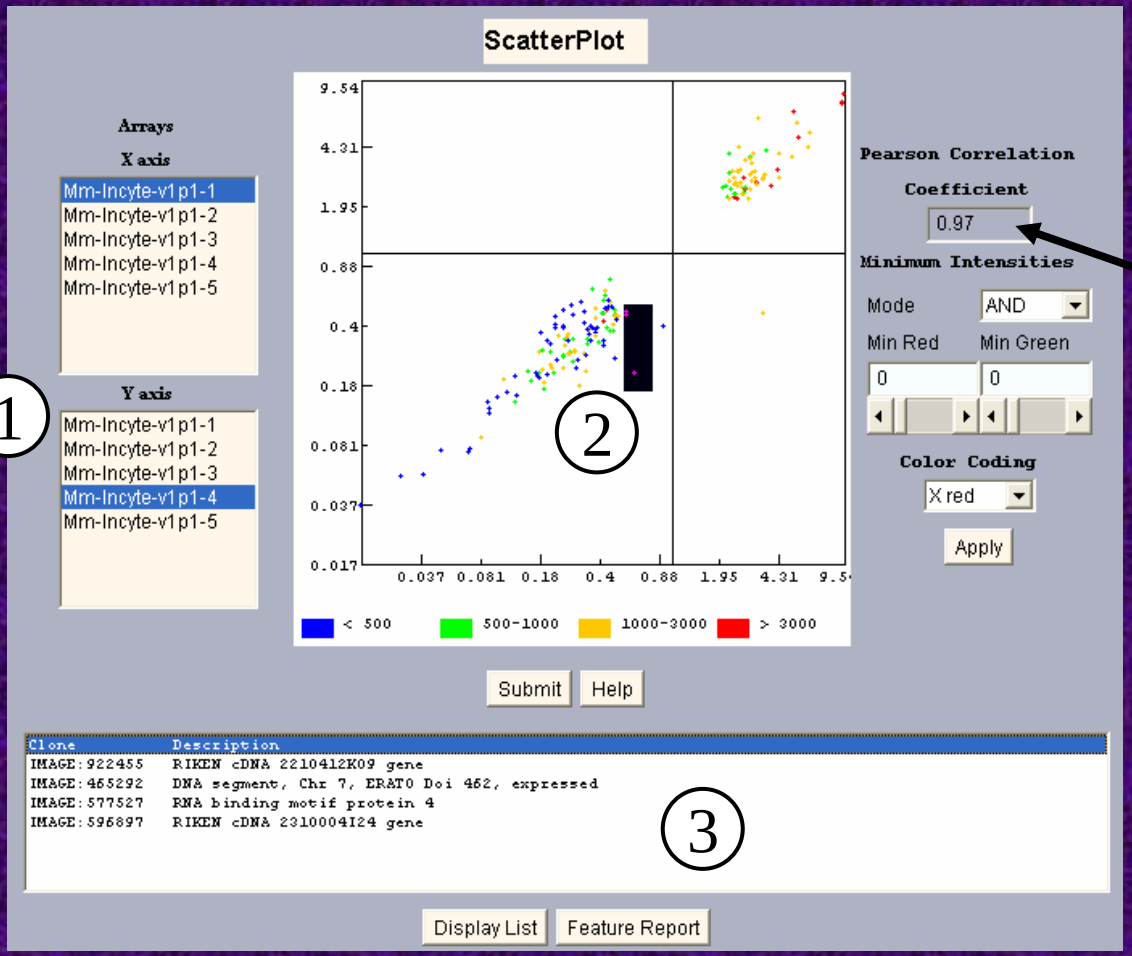
From the mAdb Gateway page, use the "Manage Identifier list" link for existing feature lists. Click on list name to view/edit.

Evaluating correlation between two arrays



From the mAdb Dataset Display Page, select the "Scatter Plot – log Ratios" Tool and hit the "View" button

Visualization Tools – Interactive Scatter Plot Applet



- Replicate experiments should be on a 45° angle (slope of 1) and the Pearson Correlation Coefficient should be approaching 1

- Reverse fluor experiments should have a Pearson Correlation Coefficient approaching -1

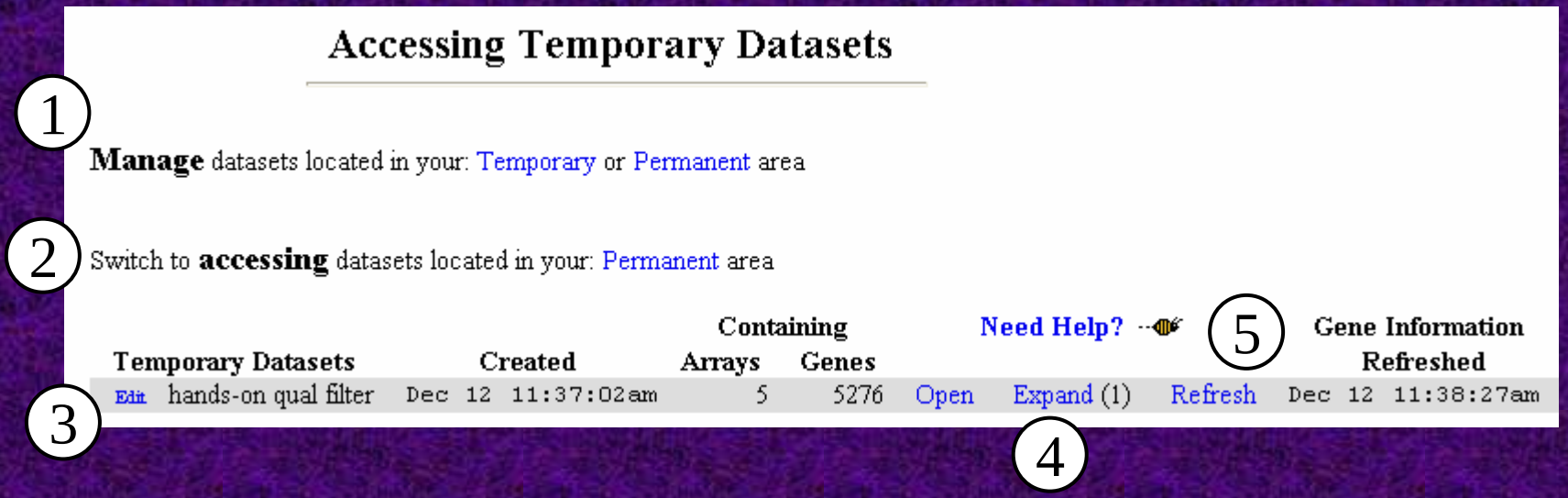
Access from *Interactive Graphical Viewers* Menu on main **mAdb Dataset Display** page:

1. Choose Arrays to be compared on X and Y axes
2. Can select outlying spots with mouse – genes will be shown in window below plot
3. Can get **Feature Report** by clicking on gene name in lower display box

V. Managing your data

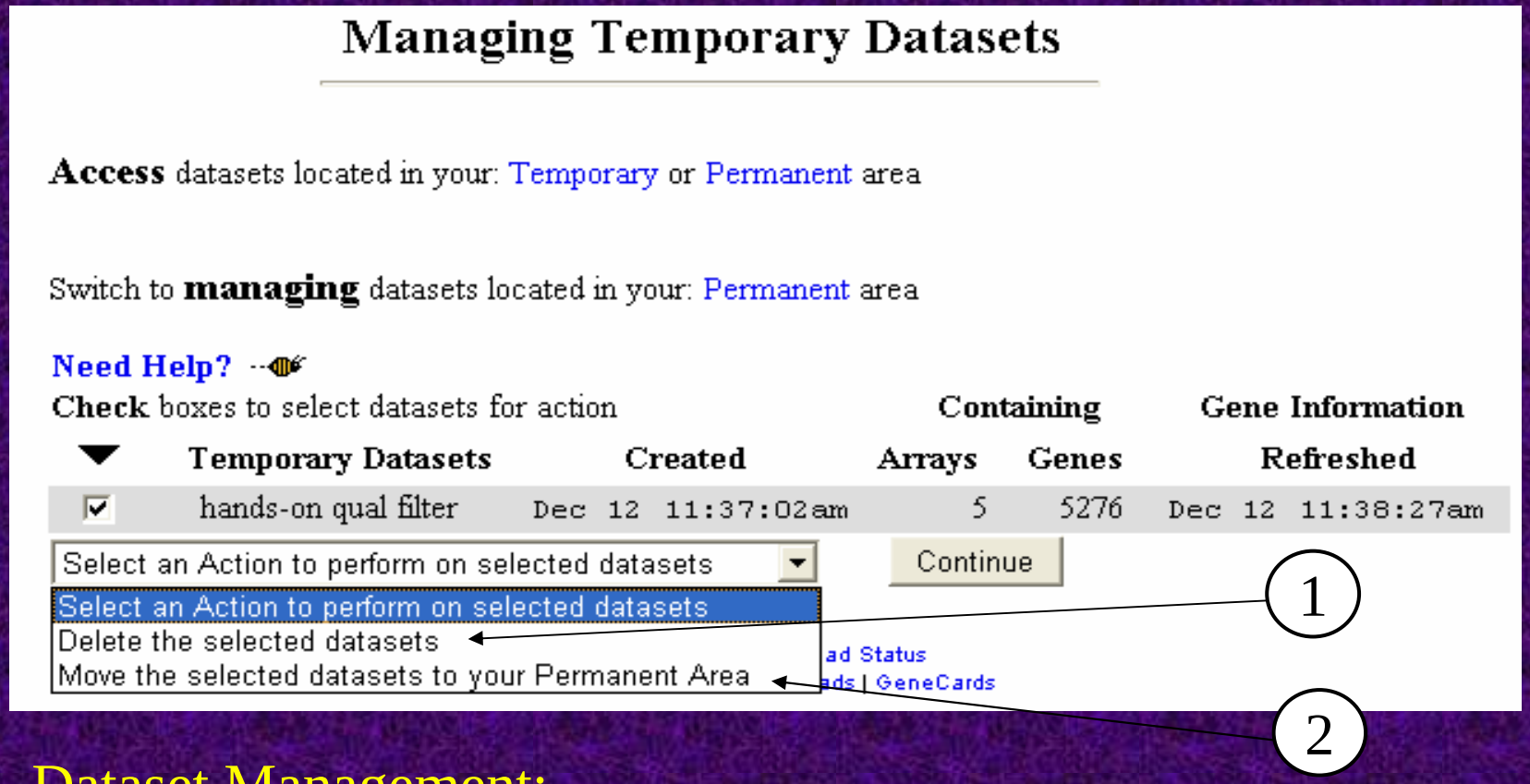
Lab 4 – Dataset Management

Goal: To keep track of your analyses and share them with others.



Dataset Access Links:

1. **Manage <u>Transient</u>**, **<u>Temporary</u>**, or **<u>Permanent</u>** Areas
2. **Access** other dataset areas which contain data (i.e. **<u>Permanent</u>**)
3. **Edit** dataset name
4. **Expand** to see parent dataset and all children of that parent
5. **Refresh** Gene Information



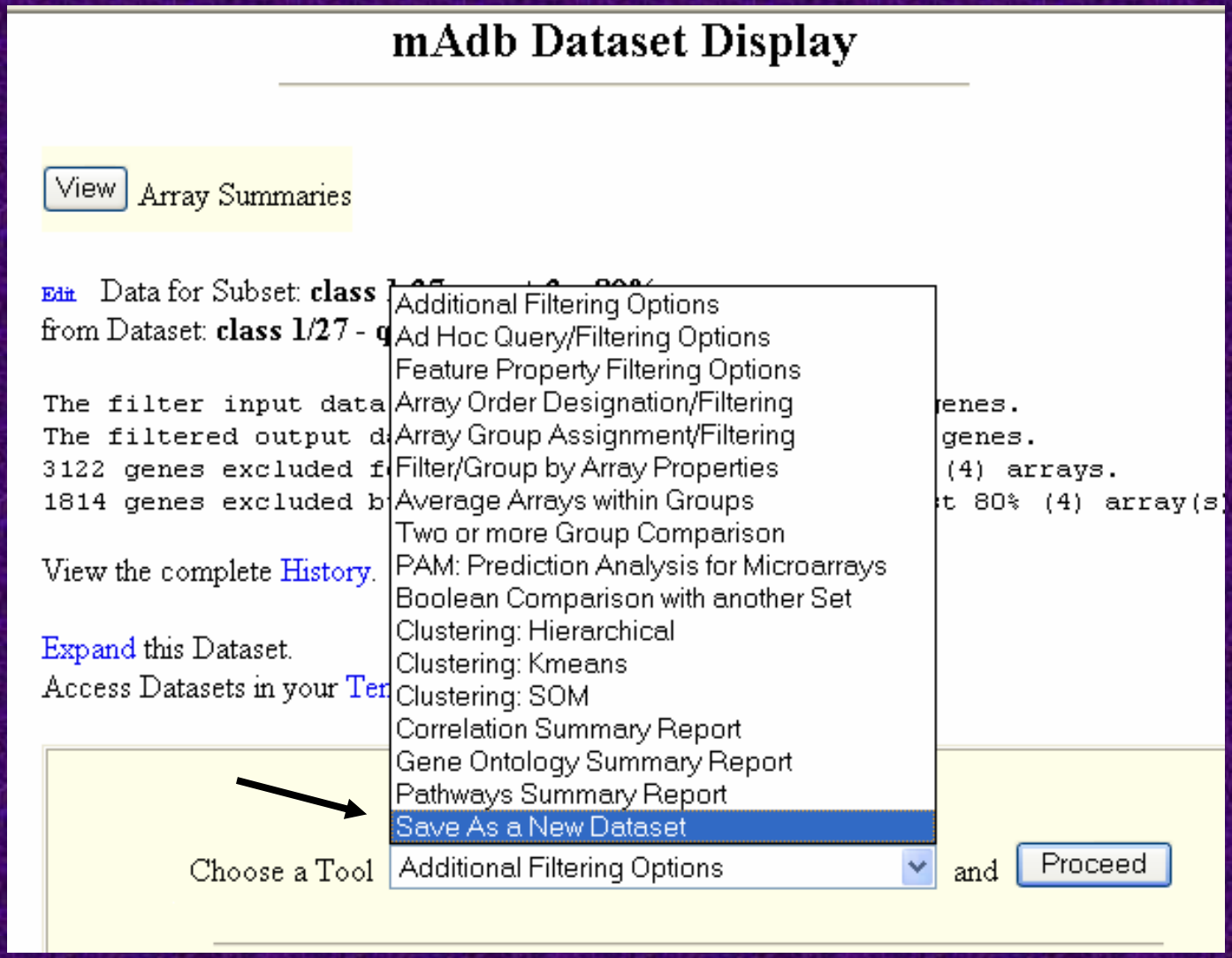
Dataset Management:

1. Can delete a dataset – but must delete parent and all children!
2. Can promote datasets (Transient to Temporary or Permanent; Temporary to Permanent)

Updating Dataset Gene Information

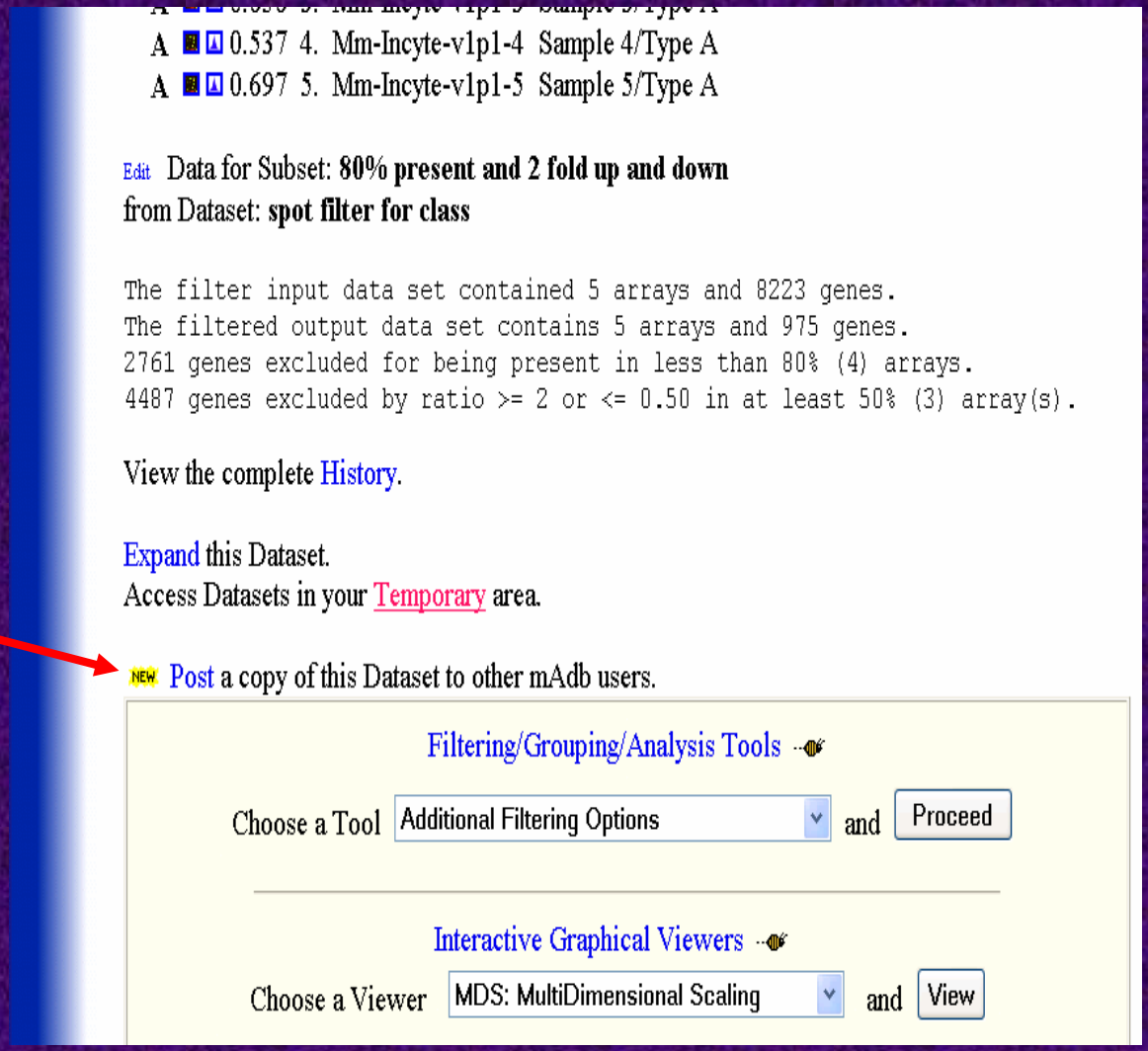
- Clicking the “refresh” link updates all of the gene information in the dataset (UniGene cluster, Description, Pathway info, Map info…)
- May want to “Save as a New Dataset”, and then refresh, if you want to keep previous annotation information

Save as New Dataset



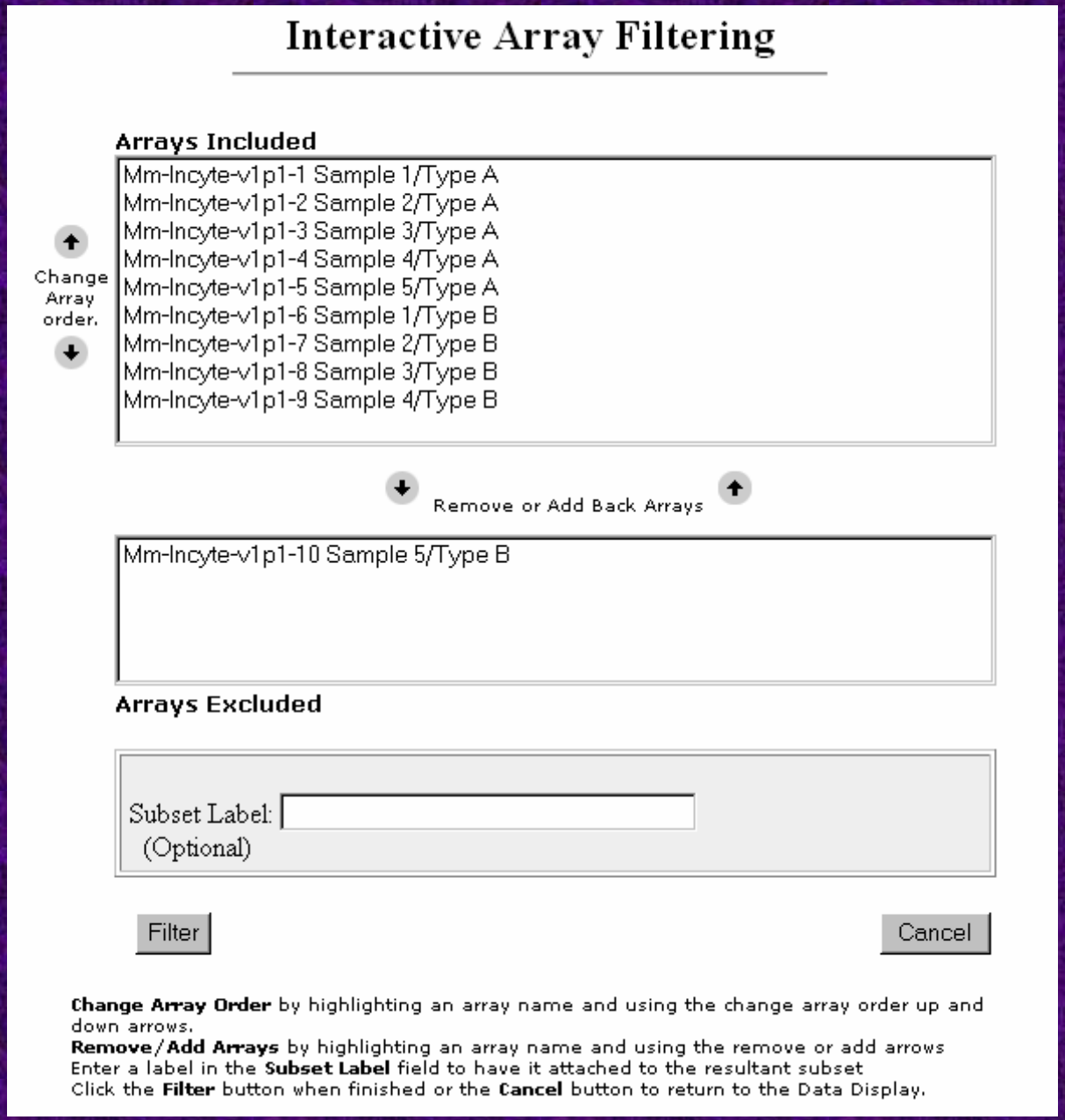
At any time, researchers can save a subset as a new dataset
In effect, this starts the tree of subsets over again at the top…

Sharing a Dataset



At any time, researchers can place a snapshot of their entire dataset including their analysis steps to other users.

From the mAdb Dataset Display Page, click on the “Post” link

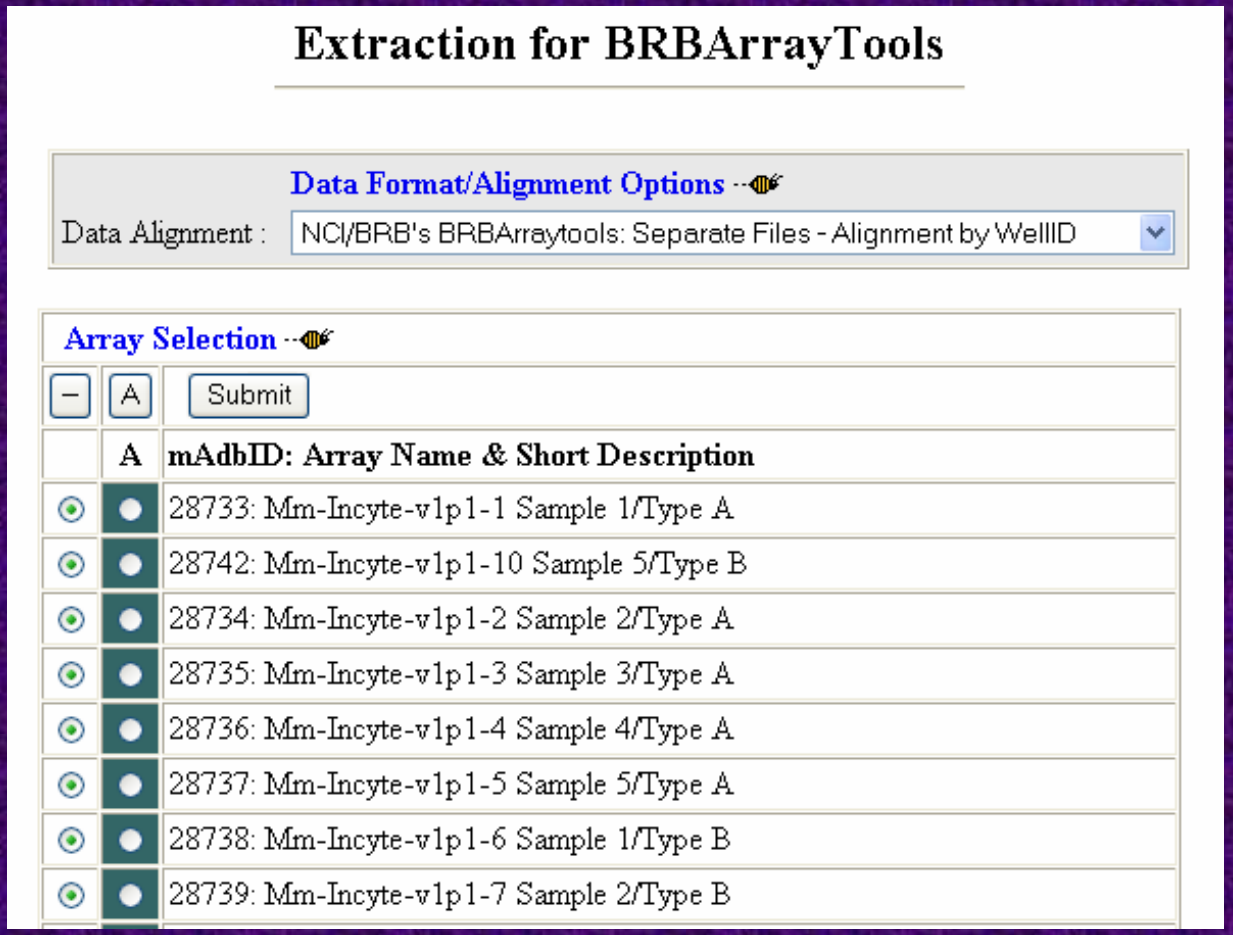


Allows re-ordering and removal of arrays from a subset

From the mAdb Dataset Display page, select the "Array Order Designation / Filtering" Tool and hit the "Proceed" button.

Exporting Data to Other Microarray Analysis Tools

- BRB Array tools export by well ID or by UniGene ID
- GeneSpring export
- MA Explorer export



Extraction for BRBArrayTools

Data Format/Alignment Options

Data Alignment : NCI/BRB's BRBArraytools: Separate Files - Alignment by WellID

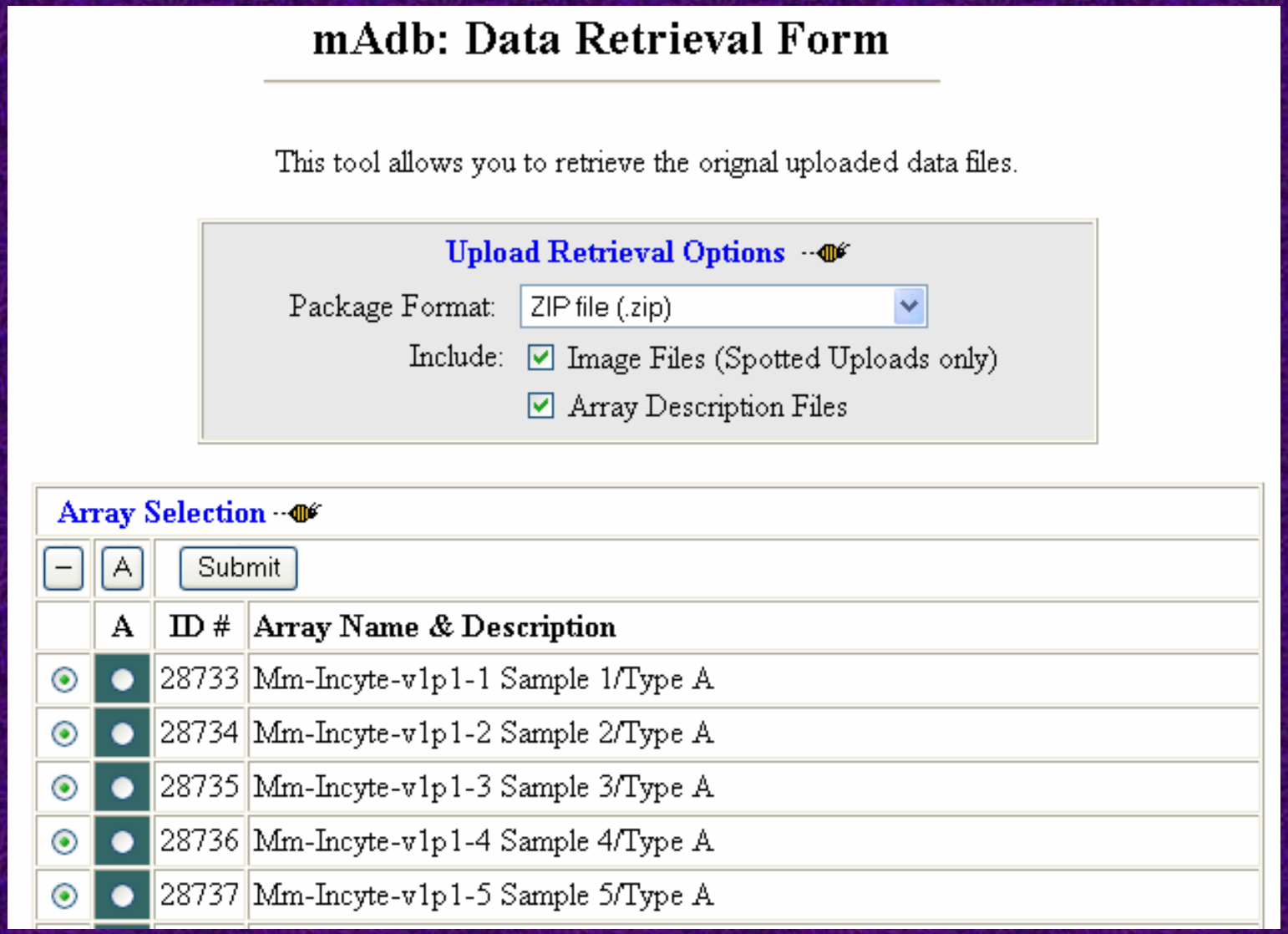
Array Selection

− A Submit

	A	mAdbID: Array Name & Short Description
◉	●	28733: Mm-Incyte-v1p1-1 Sample 1/Type A
◉	●	28742: Mm-Incyte-v1p1-10 Sample 5/Type B
◉	●	28734: Mm-Incyte-v1p1-2 Sample 2/Type A
◉	●	28735: Mm-Incyte-v1p1-3 Sample 3/Type A
◉	●	28736: Mm-Incyte-v1p1-4 Sample 4/Type A
◉	●	28737: Mm-Incyte-v1p1-5 Sample 5/Type A
◉	●	28738: Mm-Incyte-v1p1-6 Sample 1/Type B
◉	●	28739: Mm-Incyte-v1p1-7 Sample 2/Type B

From the mAdb Gateway page, select a project(s) and the "BRBArraytools Format Retrieval" Tool and hit "Continue"

Retrieving Uploaded Data



mAdb: Data Retrieval Form

This tool allows you to retrieve the orignal uploaded data files.

Upload Retrieval Options

Package Format: ZIP file (.zip)

Include: ☑ Image Files (Spotted Uploads only)

☑ Array Description Files

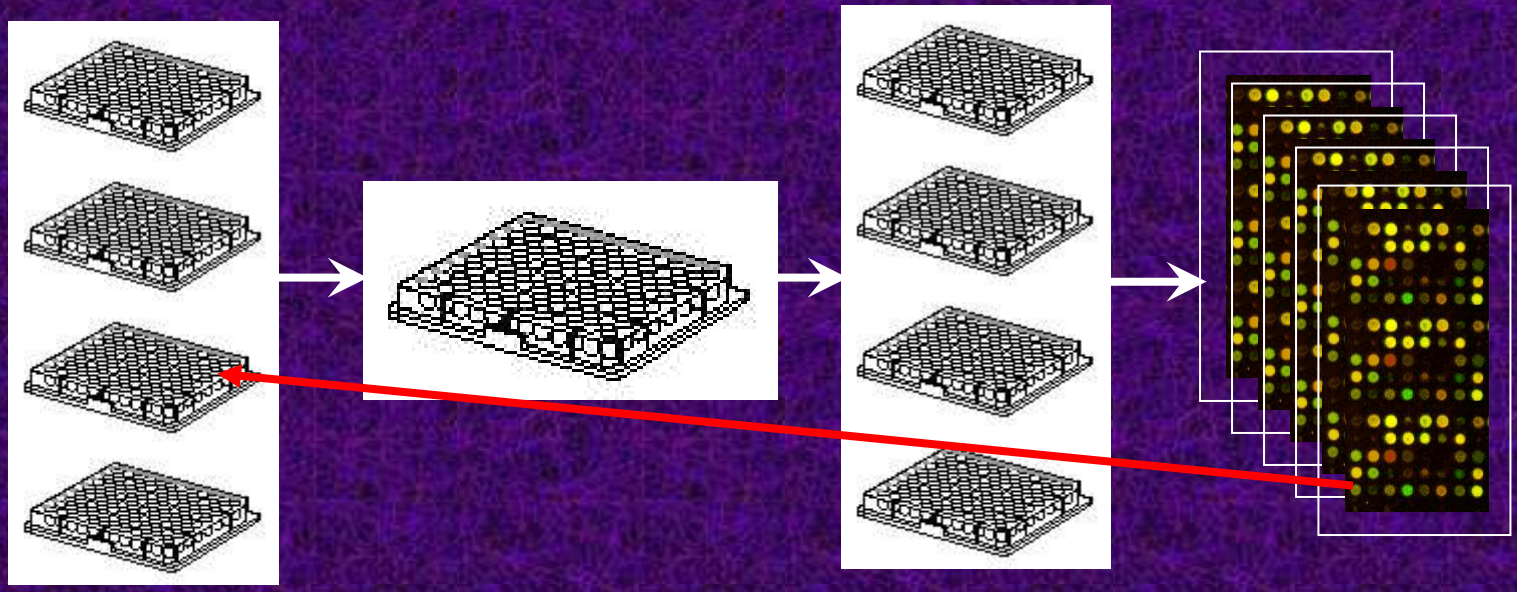
Array Selection

− | A | Submit

	A	ID #	Array Name & Description
◉	○	28733	Mm-Incyte-v1p1-1 Sample 1/Type A
◉	○	28734	Mm-Incyte-v1p1-2 Sample 2/Type A
◉	○	28735	Mm-Incyte-v1p1-3 Sample 3/Type A
◉	○	28736	Mm-Incyte-v1p1-4 Sample 4/Type A
◉	○	28737	Mm-Incyte-v1p1-5 Sample 5/Type A

From the mAdb Gateway page, select a project(s) and the “Uploaded Files Retrieval” Tool and hit “Continue”

mAdb Database Design: Feature Tracking



- mAdb works with microarray facilities to track printing from arrays back to inventory plates
- Allows mAdb support staff to correct printing errors in the database

Review of Basic Data Analysis Tools

- Within an extracted dataset, you can:
  - Filter for missing values and/or gene ratio levels
  - Do an *ad hoc* Keyword search
  - Filter datasets by lists of gene identifiers
  - View GO and Pathway Summaries
  - View data graphically
    - Interactive Scatter Plot
    - Correlation Summary Report
    - Multiple Array Viewer

mAdb Tips for array analysis

- Always look at Project Summaries – normalization factor for a "good" array should be between 0.5 and 2.0.
- If you have replicate arrays (and you should), do a scatter plot or correlation summary report to determine the correlation between the arrays (i.e. how close the slope is to 1. For reverse fluors, how close to –1) just for QC purposes.
- Turning **Show Spot Images** off, generally displays results faster – only need for spot QA.

General tips for array analysis

At a recent Microarray Data Analysis conference in Washington D.C., several speakers laid out what distinguishes a good microarray experiment from a bad one:

- When possible, consult a statistician before you even design your experiment - they offer more than just analysis tools.
- Do a power analysis to determine the number of replicates (i.e. chips) you need to detect an effect. To estimate the effect size, you might want to run a pilot study first or obtain the estimate from previous similar experiments. Regardless of the power analysis results, obtain at least three replicates on different slides or chips.
- Find sources of technical variation before you embark on a hunt for biological effects and standardize your protocols.
- Randomize your variables: for example, don't run all your treatment slides on one day and all your controls on the next.
- Microarray analysis is a screening tool – confirm your observation by other methods – RT-PCR, Northern blot, protein levels
- See **http://linus.nci.nih.gov/~brb/TechReport.htm** for good references on design, analysis issues, and myths/truths

Other microarray training

- Hands-on analysis tool mAdb class #412 – next class May 23-24
- Statistical Analysis of Microarray Data & BRB Array Tools (from the NCI Biometrics Research Branch) class #410 offered bimonthly; next class April 26-27
- Partek Pro, R, GeneSpring classes – http://training.cit.nih.gov
- Microarray Interest Group
  - 1st Wed. Seminar, 1st and 3rd Thu. Journal Club
  - To sign up: http://list.nih.gov/archives/microarray-user-l.html
- Class slides available on "Reference" page
- Sample datasets to try out the system are available from a link on the Gateway Page

mAdb Development and Support Team:

- John Powell, Chief, BIMAS, CIT
- Jim Tomlin
- Lynn Young, Ph.D.
- Esther Asaki*
- Yiwen He, Ph.D.*
- Kathleen Meyer*

*SRA International contractor

NIAID

CIT

http://madb.nci.nih.gov
http://madb.niaid.nih.gov

For assistance, remember:

madb_support@bimas.cit.nih.gov

Thank you!!